Routledge
Taylor & Francis Group

Journal of Cognitive Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pecp21

Auditory distraction of vocal-motor behaviour by different components of song: testing an interference-by-process account

Rona D. Linklater, Jeannie Judge, Patrik Sörqvist & John E. Marsh

To cite this article: Rona D. Linklater, Jeannie Judge, Patrik Sörqvist & John E. Marsh (06 Dec 2023): Auditory distraction of vocal-motor behaviour by different components of song: testing an interference-by-process account, Journal of Cognitive Psychology, DOI: 10.1080/20445911.2023.2284404

To link to this article: https://doi.org/10.1080/20445911.2023.2284404

Published online: 06 Dec 2023.

JOURNAL OF COGNITIVE PSYCHOLOGY
https://doi.org/10.1080/20445911.2023.2284404

Routledge
Taylor & Francis Group

Auditory distraction of vocal-motor behaviour by different components of song: testing an interference-by-process account

Rona D. Linklater[a], Jeannie Judge[a], Patrik Sörqvist[b,c] and John E. Marsh [a,b]

[a]School of Psychology and Humanities, University of Central Lancashire, Preston, UK; [b]Department of Health, Learning and Technology, Luleå University of Technology, Luleå, Sweden; [c]Department of Building Engineering, Energy Systems and Sustainability Science, University of Gävle, Gävle, Sweden

ABSTRACT

The process-oriented account of auditory distraction suggests that task-disruption is a consequence of the joint action of task- and sound-related processes. Here, four experiments put this view to the test by examining the extent to which to-be-ignored melodies (with or without lyrics) influence vocal-motor processing. Using song retrieval tasks (i.e., reproduction of melodies or lyrics from long-term memory), the results revealed a pattern of disruption that was consistent with an interference-by-process view: disruption depended jointly on the nature of the vocal-motor retrieval (e.g., melody retrieval via humming *vs.* spoken lyrics) and the characteristics of the sound (whether it contained lyrics and was familiar to the participants). Furthermore, the sound properties, influential in disrupting song reproduction, were not influential for disrupting visual-verbal short-term memory—a task that is arguably underpinned by non-semantic vocal-motor planning processes. Generally, these results cohere better with the process-oriented view, in comparison with competing accounts (e.g., interference-by-content).

ARTICLE HISTORY
Received 27 April 2023
Accepted 10 November 2023

KEYWORDS
music performance; vocal motor-planning; auditory distraction; interference-by-process

Extraneous task-irrelevant sound is typically omnipresent. Humans are passive recipients of music, speech, and environmental sound around them and may even be unaware of its presence. However, through the unique, sentinel quality of the human hearing system, our auditory environments are nonetheless processed and therefore carry the potential to influence cognition. Thus, irrespective of how we are otherwise engaged, an irrelevant auditory stimulus can detrimentally impact our goal-directed behaviour. Selective attention must be permeable to enable a necessary response irrespective of current behaviour—for example, to alert danger. However, it is this quality that leaves an organism open to the necessary consequence of distraction (Hughes & Jones, 2003; Johnston & Strayer, 2001). Previous exposure to sound, such as listening to music before conducting a task, can sometimes enhance cognitive task performance—an effect attributed to the benefit of increased arousal or positive mood (Perham & Whitney, 2012; Schellenberg, 2005; Thompson et al., 2001). However, a more common consequence of the passive processing of concurrent sound while engaged in mental activity, is the disruption of cognitive task performance, compared to quiet (e.g. Colle & Welsh, 1976; Furnham & Strbac, 2002; Perham et al., 2013; Thompson et al., 2011). For example, reading comprehension is impaired by background speech (Martin et al., 1988; Sörqvist et al., 2010) and to a greater extent by the presence of lyrical, than non-lyrical music (Martin et al., 1988; Perham & Currie, 2014). In the current study, we explore whether the disruptive effects of task-irrelevant sound, especially different components of song, are jointly characterised by properties of the sound and properties of the mental processes underpinning a focal task (cf. Jones & Tremblay,

CONTACT John E. Marsh jemarsh@uclan.ac.uk Human Factors Laboratory, School of Psychology and Humanities, Darwin Building, University of Central Lancashire, Marsh Lane, Preston. PR21AE, UK

2000). Specifically, we address whether the characteristics of sound that disrupt tasks involving semantic memory for music and lexical retrieval, differ from those that require visual-verbal short-term memory.

The empirical platform upon which most work on auditory distraction has been undertaken uses a visual-verbal short-term memory task (e.g. Beaman & Jones, 1997; Conrad, 1964). This task requires recall of to-be-remembered information (usually digits or letters) in the exact order of its initial presentation. An appreciable decrease in serial short-term memory performance in the presence of to-be-ignored sound has been termed the "irrelevant sound effect" (ISE, Beaman & Jones, 1997; see e.g. Colle & Welsh, 1976; Hughes et al., 2007; Jones & Macken, 1993; LeCompte, 1996; Salamé & Baddeley, 1982). Several accounts have been put forward to explain the ISE and one purpose of the current study was to augment theoretical debates regarding mechanisms of auditory distraction (e.g. Bell et al., 2019; Hughes, 2014).

Theories of auditory distraction

The classical structuralist view (e.g. Baddeley, 1986) promotes the idea that focal task disruption from irrelevant sound is due to interference caused by the structural similarity between to-be-remembered and to-be-ignored items that cohabit a short-term store or space. Identified as "interference-by-content" accounts, they have numerous manifestations (Neath, 2000; Oberauer & Lange, 2008; Salamé & Baddeley, 1982). In the context of the ISE, for example, this class of accounts suggest that background speech impairs the memory of lists of stimuli because the to-be-ignored stimuli (the speech) and the to-be-recalled items are similar in content.

An alternative explanation, identified as "interference-by-process" (Jones et al., 1996; Jones & Tremblay, 2000), proposes task impairment from auditory distraction arises due to concurrent, common, mental processes operating during the serial organisation of stimuli. According to this account, disruption produced by irrelevant sound on visual-verbal serial recall is attributable to a clash between two contemporaneous serial-ordering mechanisms: one deliberate and applied to the focal task—serial rehearsal—and one automatic and applied pre-attentively to the irrelevant sound (Jones et al., 1996; Tremblay & Jones, 1998). Seriation processes must apply for both focal task and irrelevant sound processing for disruption to take place. Disruption therefore happens irrespective of the content of the sound (e.g. whether it comprises language or not) and irrespective of whether it is familiar or not (Jones et al., 1990). In this view (e.g. Jones & Tremblay, 2000), any disruption to focal task performance by irrelevant sound will be determined by factors related to perceptual streaming (see Bregman, 1990).

Changes in the acoustic make-up of the stimulusare important determinants of disruption. Changing speech letters (e.g. d, w, r, l or varied pitch, A, C, F#, B), are more distracting relative to repeated sounds (e.g. d, d, d, d or constant pitch A, A, A, A). This is called "the changing state effect" (Jones, 1993) and plays a central role in the interference-by-process account (Jones & Tremblay, 2000). A series of changing-state speech or non-speech sounds will invariably produce greater disruption of serial recall than a sequence of unvarying material (Jones & Macken, 1993). The fact that the changing-state effect is determined by the sounds' acoustic properties, rather than the sound items' identity or content, aligns with the interference-by-process view but contradicts that of interference-by-content (Neath, 2000; Salamé & Baddeley, 1982).

The concept of interference-by-process is encompassed within an emerging account of short-term memory phenomena that has been coined the perceptual-gestural account (e.g. Jones et al., 2006), or perceptual-motor view (e.g. Hughes et al., 2009; Hughes & Marsh, 2017; Jones et al., 2004). This view assumes that short-term memory performance is parasitic on processes and systems that are general-purpose mechanisms not specifically dedicated to memory. The perceptual-motor view (Hughes & Marsh, 2017; Jones et al., 2004, 2006) eschews the notion that inner speech (sub-vocal rehearsal) serves the function of refreshing decaying memory traces within a mnemonic store or space, as promoted, for example, by the phonological store interference-by-content account (e.g. Baddeley, 1986, 2007). Rather, the skill of speaking (overt or inner-speech) is exploited to bind together items—that have no pre-existing syntactical or grammatical cues as to their order—into a single temporally extended motor-plan on a common carrier. Speech, being necessarily sequential, is thus well suited to embodying the serial order of items and co-articulatory and prosodic

characteristics of natural speech can further imbue the motor-plan with serial order cues (Macken et al., 2014; Woodward et al., 2008). Furthermore, forward internal models that operate to enable sensory outcome predictions of planned or imagined actions (sensory, auditory or visual) can influence the online processing of external sensory information (e.g. music) within the environment (Halász & Cunnington, 2012; Maes & Leman, 2013; Schutz-Bosbach & Prinz, 2007; Witt, 2011). This suggests an intricate relationship between motor planning and musical perception, similar to the relationship between perceptual organisation and motor planning (e.g. Hughes et al., 2016). In support of this relationship research suggests that long-term experiences (and activities) involved in playing a musical instrument yield a superior association between sensory information and fine motor movements in musicians, as compared to non-musicians, that gives rise to superior sequence learning (e.g. Anaya et al., 2017).

Vocal production

The perceptual-motor view explains the greater disruption of serial recall observed from changing-state (as compared to steady-state) sound sequences as being due to the motor-plan's susceptibility to the obligatory process involved in perceptually organising sound into streams (Hughes & Jones, 2005; Jones & Macken, 1993). Providing successive sounds within a changing-state sequence, verbal or non-verbal, are acoustically similar to one another (in frequency or timbre) and share a common ground (e.g. voice), they are assigned to the same stream. Their serial order sequence is formed via the assignment of successive cues to the same stream, which results in it being a candidate for populating the motor-plan (Hughes, 2014; Hughes & Jones, 2005; Hughes & Marsh, 2017; Jones et al., 2004; see also Marsh et al., 2009). Therefore, according to the perceptual-motor view, limitations in performance (in the context of serial recall) by changing-state sound are not due to the limitation of mnemonic capacity within a putative static short-term memory store or space that is determined by the existence of items prone to decay, interference, or displacement as interference-by-content accounts hold (Baddeley, 1986; Neath, 2000). Rather, it represents a competition-for-action, whereby disruption to serial recall is produced by the conflict of changing-state between the target items and the irrelevant sounds (Hughes, 2014). While the interference-by-content accounts focus on the individuality of each item presented (and their rate of decay, or structure, such as phonological similarity; Baddeley, 1986; Nairne, 1990; Neath, 2000), the perceptual-motor view focuses on sequence factors rather than constituent, individual items, and means of output via motor-planning that enable sequences to be assembled and sub-vocally rehearsed (Hughes & Jones, 2005; Jones et al., 2006, 2007).

The generality of the vocal-motor disruption via background sound, as suggested by the perceptual-motor view (e.g. Hughes & Jones, 2005), implies that it should not be confined to short-term serial recall. It should also extend to other tasks involving motor-planning and sequential verbal production. For example, the automatic processing of irrelevant auditory input could assume/threaten to assume control of the sub-vocal motor system responsible for the planning of, and production of, vocalisable components of song (cf. Godøy & Leman, 2009; Leman, 2007; Leman & Maes, 2014, 2015). One goal of the current study was to determine whether the perceptual-motor view (Hughes & Jones, 2005; Hughes & Marsh, 2017) and the construct of interference-by-process, could be extended to a hitherto unstudied research domain involving vocal retrieval of known song components (melody, lyrics) from long-term memory.

Later in this empirical series, we introduce accounts that attribute auditory distraction to an attentional capture mechanism. Specifically, it will be investigated whether some of the disruptive effects of to-be-ignored sound reported in the current investigation align better with an account that assumes that particular properties of sound are capable of wresting attention from a focal task, thereby resulting in performance decrement (Bell et al., 2010, 2012; Hughes et al., 2005, 2007; Marsh et al. 2018; Parmentier et al., 2018; Röer et al., 2013; Vachon et al., 2017).

A considerable body of existing research has focused on the impact of presenting irrelevant sound/music during short-term memory recall/recognition tasks of visually presented notated tones/words. By contrast, little research focusses on the impact of irrelevant sound/music on vocal production. It is conceivable that task-irrelevant sound, known to disrupt the vocal-motor planning process in short-term memory (Hughes & Marsh,

2017), may impact on the vocal production mechanisms required for melody/lyrics retrieval from long-term memory. Existing studies have focussed more on exploring music representation within memory, rather than its production (e.g. Fiveash et al., 2018; Miranda & Ullman, 2007; Perruchet & Poulin-Charronnat, 2013; Thompson & Yankeelov, 2012).). Consequently, it is unclear how processes involved in vocal input/output are related to one another and, of greater consequence, how vocal-motor planning required for vocal music/language production may be affected by a competing motor-plan provided by the presence of extraneous music or language. One aim of the current study was thus to explore the properties of background sounds that impact on the vocal-motor planning process in the context of a familiar song.

Watkins and Allender (1987, p. 565) state: "it is surely more difficult to bring to mind a particular tune if another is being heard". The motivation for the current work was driven, in part, by the lack of any known empirical study exploring this anecdote and ensuing questions to which it gives rise: Is it, in fact, more difficult to produce a familiar song when hearing another? If so, which features of the background music (e.g. melody, lyrics, familiarity) are responsible for promoting the disruption? Does the nature of the production of song (e.g. speaking lyrics *vs*. humming melody) also interact with the features of the background sound to determine disruption? Addressing such questions could potentially inform the nature of the perceptual and cognitive processes underpinning memory for, and production of, song, and more generally inform the nature of verbal behaviour and memory functions at large.

Integration versus independence

Language and melody within song are thought to be subtended by independent, hierarchical systems that have comparable properties (Schön et al., 2004). For example, discrete items (notes of pitch, words) are combined into musical chords, or sentences, with syntax rules governing the operation (Fiveash et al., 2018; Thaut, 2009). A continued source of debate, however, is whether the language and melody components of song are integrated or processed independently when listening to song (Hamzelou, 2010; Hébert & Peretz, 2001).

The finding that melodies of songs are better recognised when heard with their original words than when heard with text of another equally familiar song, and vice versa (Crowder et al., 1990; Samson & Zatorre, 1991), is taken as support for the integration of melody and lyrics. However, some neurological studies have demonstrated independent processing of melodic and semantic song components (e.g. ERP[1] studies, Besson et al., 1998; PET[2] studies, Gröussard et al., 2010; see also Bonnel et al., 2001).

Exploring the impact of different properties of task-irrelevant auditory distracters on the production of melody (Experiment 1) and lyrics (Experiment 2) in the present study, affords a window onto the nature of processing within the mnemonic systems responsible for the two properties of song, thereby shedding light on the integration *versus* independence debate. According to the integration view (Schön et al., 2010; Serafine et al., 1986), we might expect the pattern of any disruption attributable to task-irrelevant sound with different properties, to be observed regardless of whether melody production (e.g. humming: Experiment 1) or lyrics production (Experiment 2) is required. On the flipside, a deviation to the patterns of interference observed could be taken as *prima facia* evidence for independence (e.g. Besson et al., 1998; Bonnel et al., 2001). As mentioned in the foregoing, supposing that the presence of background music does indeed impair vocal production of melody or lyrics, it is important to determine whether an effect is underpinned by prior memory for, and thus familiarity with, distracter stimuli.

Influence of distracter familiarity

To understand the potential influence of task-irrelevant music on the production of familiar melody (via humming; Experiment 1) or lyrics (via speaking; Experiment 2) one might look to modular models of music processing (e.g. Peretz & Coltheart, 2003). These models propose the presence of brain specialisation, a "modular architecture", for local neural circuitries essential for music processing (Peretz & Coltheart, 2003, p. 689). Within the

[1]Event related potential (ERP).
[2]Positive emission tomography scan (PET).

context of this modular model, memory for familiar melodies, but not necessarily the temporal process of learning, is generally understood to denote musical semantic memory.

The embodied music cognition theory suggests music perception activates motor systems that allow melodies to be reproduced or performed (Leman, 2007; Leman & Maes, 2015; Sievers et al., 2013). Therefore, the production of a familiar target melody (with associated lyrics) or lyrics from long-term memory and the mere perception of a different familiar to-be-ignored melody/lyrics should involve some degree of activation within the motor system. For example, subvocalisation (covert singing of lyrics using inner speech) has been implicated in the rehearsal of auditory images (the imagining of an auditory stimulus [e.g. music] without hearing the actual sound, utilising the inner ear, Halpern, 2001; Pring & Walker, 1994; Reisberg et al., 1989; Smith et al., 1995). Moreover, since familiar, as compared to unfamiliar, stimuli are well-represented in long-term memory, such vocal-articulatory activation should be greater for familiar than for unfamiliar stimuli and should, therefore, be more disruptive to the vocal production of another song. The current study set out to test these predictions.

Experiment 1

The incentive for Experiments 1 and 2 of the current study was guided by the need to explore whether and how retrieval processes for song operate and interact. The vocal production mechanism, responsible for producing components of familiar song (e.g. Peretz & Coltheart, 2003; Tsang et al., 2011), was explored by investigating its propensity to be disrupted by the mere presence of familiar and unfamiliar task-irrelevant music and/or language. In Experiment 1, participants hummed the melody of familiar songs that were cued by title. Humming was chosen because, arguably, it does not necessitate access to the phonological lexicon (Peretz & Coltheart, 2003) but nonetheless requires low-level sensorimotor operations of singing mechanisms such as pitch control and timing of movement (Özdemir et al., 2006; Schubotz et al., 2000). While humming target melodies in Experiment 1, participants were exposed to to-be-ignored sounds comprising: (1) a different instrumental melody that was familiar or unfamiliar to participants, (2) sung familiar or unfamiliar lyrics to a familiar or unfamiliar melody, and (3) spoken lyrics from different familiar or unfamiliar song. Inclusion of these conditions made it possible to study characteristics of the features of song that drive disruption of vocal motor-processing by sound.

The design of Experiment 1 enables us to address the suggestion that familiar songs are more difficult to produce when hearing another (Watkins & Allender, 1897). It allows insight into which features of background music (melody, lyrics, familiarity) promote disruption. In turn, the design can also inform whether language and music are integrated or independent within song (Besson & Schön, 2001; Peretz & Zatorre, 2005; Thompson & Yankeelov, 2012).). From the standpoint that verbal retrieval is vulnerable to distraction from activation of similar information stored within semantic memory (Jones et al., 2012) it was hypothesised that production of a familiar melody would be more impaired by the presence of a task-irrelevant familiar melody—that has a representation within long-term memory and can therefore compete with the target melody for retrieval—than an unfamiliar melody combining a novel pitch contour with a familiar rhythmic pattern. However, it was also hypothesised that unfamiliar melody, acting as an irrelevant auditory distraction, would produce some disruption as compared with quiet (Jones & Macken, 1993; Salamé & Baddeley, 1989). If humming does not require access to the phonological lexicon (cf. Peretz & Coltheart, 2003), then it was expected that an irrelevant familiar melody with sung lyrics would produce no more disruption than a familiar melody without lyrics. However, if access to the phonological lexicon was required, a distracter with lyrics would produce a greater disruption than without lyrics. Moreover, if stored separately, spoken lyrics (familiar or unfamiliar) without melody should be less disruptive than spoken lyrics with a familiar melody.

Method

Participants

Two hundred and ninety-four adults aged 18–92 (mode range 45–49[3]), from local community/university groups participated, in return for travel reimbursement or course credits. As multiple between-

[3]Participants ticked an age-range, so no exact ages were obtained.

participants conditions were used, to ensure comparability between the participants across groups, all 150 participants aged over 50 years completed the Addenbrooke's Cognitive Examination Revised (ACE-R[4]). A mixed analysis of variance (ANOVA) showed Sound Condition × Cognition to be non-significant.[5] Visual acuity and hearing tests confirmed acceptable levels, with Ethical Clearance for the ensuing four experiments obtained from the University of Central Lancashire. To determine an appropriate sample size in each experimental condition, we conducted two apriori power analyses (using G*Power; Faul et al., 2007). There are, to the best of our knowledge, no previously published experiment that is identical or nearly identical to the experiments conducted in the present paper. Because of this, we based the power analyses on the theoretical assumption that the effects of the present study would resemble the effects of background sound on retrieval from long-term memory reported in Marsh, Crawford, et al. (2017) and in Jones et al. (2012, Experiment 3). The effects reported in their studies are conceptually similar to the effects under study in the current investigation. The experimental effect (the difference between the two Sound Conditions) in Marsh, Crawford et al.'s study had an effect size of $d_z = 0.41$. The estimated sample size needed to detect an effect (one tailed) of this size is 66 participants. Moreover, the experimental effect (from a repeated measures analysis of variance across three Sound Conditions) reported in Experiment 3 (that had spoken output as dependent measure, similar to the current study) in Jones et al. (2012) had an effect size of $\eta_p^2 = .22$ (from which we obtained an effect size of the F of 0.53). The estimated sample size needed to detect this experimental effect is 11 participants. Taken together, we aimed to collect data from about 70 participants in each condition, which should be enough to detect the effects according to the power analyses.

Materials and apparatus

Thirty target and ten distracter songs were compiled from western culture nursery/traditional rhymes following a pre-study questionnaire completed by 20 non-participant volunteers to assess familiarity. All songs shared distinctive intrinsic characteristics, with simple duple, triple, quadruple, or compound duple time signatures, a duration range of 12–24 s ($M = 18.4$), and a beat range within each excerpt of 24–48 ($M = 36.9$). Pace in western music is indicated by beats per minute (bpm) and each excerpt had the tempo of 120 bpm to a crotchet pulse (quarter note). Time values included quavers, crotchets, and minims, with occasional dotted quaver + semiquaver combinations. A trained contralto pre-recorded a live a cappella performance, and monotonic spoken lyrics (paced to the tempo of the sung version [Simmons-Stern et al., 2012]), into a laptop computer via a line-in USB microphone using Audacity. Instrumental stimuli were computer generated using synthesised sounds of a concert flute from the "Avid Sibelius 7" music software notation program. Phrasing was legato throughout. All stimuli were looped from their original times to run for 32-second to allow equal duration with the 32-second time requirements to complete the humming for each target melody. For the unfamiliar melodies, while tonality, rhythmic pattern, and implied harmony were matched to the original melody, the pitch contour was reformed to create novel interval progressions. Unfamiliar lyrics followed the original syntactic structure, syllabic setting with limited melisma (see Figure 1(a,b)). Irrelevant sound was presented via Sennheiser HD201 headphones, averaged at 62 dB(A), executed by the E-Prime program (2) software tool (Taylor & Marsh, 2017) via a laptop computer. The dynamic was consistently mezzo-forte *(mf)* with no tone gradation.

Design

A mixed 3 (Sound Condition) × 4 (Sound Type) design was used, and the within-participants factor of Sound Condition was classified into three levels, quiet (no distracter), familiar distracter present, and unfamiliar distracter present. Each level comprised 10 trials, and each of the four song sets (the fourth being the distracter set) contained 10 melodies. The between participants factor of Sound Type (Group) had four levels, Melody, Familiar Lyrics, Speech (spoken lyrics),

[4] $M = 92.75$, $SD = 4.28$, Mini mental state examination (MMSE) $M = 28.86$, $SD = 1.19$: no cognitive impairment.
[5] $F(2, 296) = 2.042$, $MSE = .224$, $p = .132$, $\eta_p^2 = .014$ (Based on ACE-R data).

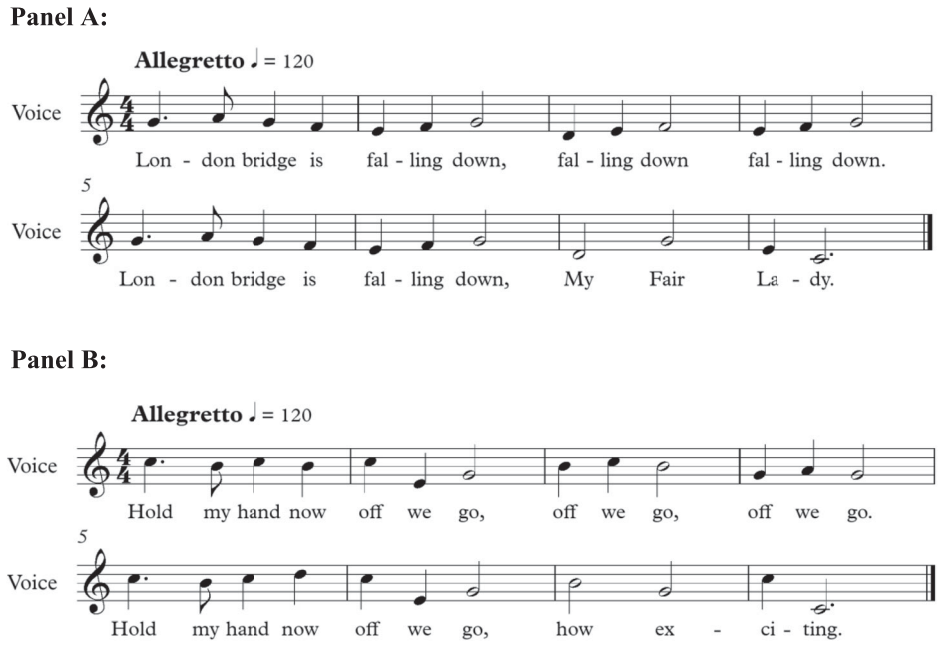


Figure 1. An example of familiar melody with familiar lyrics (panel A) and an example of unfamiliar matched melody with unfamiliar lyrics (panel B).

Table 1. Combination of experimental Sound Conditions for Sound Type Groups.

Sound Type Group	*n*	Distracter variable
All	294	quiet
Familiar Lyrics	78	familiar sung-lyrics (familiar melody)
		familiar sung-lyrics (unfamiliar matched melody)
Melody	72	familiar melody (no lyrics)
		unfamiliar matched melody (no lyrics)
Speech	72	familiar spoken-lyrics (no melody)
		unfamiliar matched spoken-lyrics (no melody)
Unfamiliar Lyrics	72	unfamiliar sung-lyrics (familiar melody)
		unfamiliar sung-lyrics (unfamiliar matched melody)

and Unfamiliar Lyrics, and participants were alternately allocated to each Sound Type (Group) according to the balance of numbers per condition. Each Sound Type had 72 participants with an even distribution of age (except 78 participants participated in the Familiar Lyrics Group) as shown in Table 1. To address repeated measure unsystematic variation, participants were alternately assigned into four 72-participant Groups,[6] the six orders of presentation within each being counterbalanced. All participants experienced a quiet control condition.

Hummed recall accuracy of melodic contour was the dependent variable, prior knowledge was established in a final Recognition test that involved indicating whether participants recognised familiar against unfamiliar material using the same stimulus type (e.g. sung-lyrics, melody, spoken-lyrics) they had previously experienced within the experiment. Although not a key goal, a mixed 3 (Sound Condition) × 4 (Sound Type) design also computed the onset time (OST) to begin each melody performance.

Procedure

Participants were tested individually in a quiet room in the presence of the researcher. Following an explanation of task requirements, instruction to ignore any auditory sequences heard through their headphones, completion of a demographic response sheet[7] and consent form, two practice trials in quiet allowed familiarisation. Task response was to hum the target melody, or as much as was known, from a 32-second visually presented title before an audible bleep signalled the next title. Performances were recorded via the computer's

[6]78 participants participated in the Familiar Lyrics Group due to a programme error.
[7]Western musical culture indicated by 98.9% of participants, 2.1% indicating Other. Prior musical training/current musical participation were non-significant factors.

internal microphone. If the song title was not known, participants were instructed to remain silent until the next song title appeared on the screen. Familiar, unfamiliar, no-sound (quiet), trial blocks followed in the order generated by the E-Prime program. Song title and distracter sound, randomised by the E-Prime program and identical for each Sound Type Group, began simultaneously. A Recognition test then individually presented all 40 song titles. Participants pressed the computer "N" button if the title melody was unfamiliar, or, if familiar, the "Y" button that alerted them to hum the melody again in quiet. Recorded audio files were assessed for melodic accuracy to a scale ranging from zero, inaccurate, to four, accurate and fluent (see Appendix 2). Recordings of raw data. WAV files from all vocal performances of retrieved target melodies were measured by the author using a scale ranging from zero, inaccurate, to four, accurate and fluent, based on established musical performance assessment criteria used by exam boards (e.g. Edexcel), known and understood by the author. The scale was designed to allow a zero score for no rewardable material, the total number of bars for each melody then being divided into four sections of equal number to match the criterion descriptors. Two independent raters, musically trained, qualified teachers, also each assessed half of each Sound Type Group for participant accuracy to the target melody following the same (understood) scoring criterion. To assess author/inter-rater reliability an intraclass correlation coefficient (ICC) was conducted and showed a high consistency between author/researcher and independent raters, Cronbach's $\alpha = .966$.[8] In addition, performances were judged recognisable in relation to the song title by two non-musically trained raters, Cronbach's $\alpha = .988$.

Statistics

Within the results sections throughout the current article, we report Cohen's *d* as a measure of effect size for pairwise comparisons and the size of these effects are interpreted as small, medium, or large using Cohen's (1988) conventions. We also report Bayes factors for all pairwise comparisons. These were computed using a Cauchy prior with a scaling factor set to 1 (Rouder et al., 2009), and we used the categorisation scheme developed by Jeffreys (1961) and updated by Lee and Wagenmakers (2013) to define the strength of evidence for the alternative hypothesis and the null hypothesis, where appropriate, to back-up null hypothesis significance testing (NHST) based inferences relating to the absence of between condition differences. Since some of the key conclusions within our study rest on the observation of null effects the Bayesian approach was adopted to provide further support for the null hypothesis (H_0). Where assumptions of sphericity are violated, the Huynh-Feldt correction is deployed and reported. We report unadjusted pairwise comparisons for multiple tests. However, when the *p* values depart from those derived from the Holm–Bonferroni sequential method (Holm, 1979) to deal with family-wise error rates, we report the adjusted *p* value in parenthesis preceded by an Asterisk after reporting the unadjusted *p* value.

Results

Retrieval accuracy

As illustrated in Figure 2, the mean proportion of correct recall was affected by the presence of sound. Data was relatively consistent across Sound Type in the quiet condition although overall performance was slightly better for the Unfamiliar Lyrics Group in comparison to the Melody Group, $MD = .203$, $SE = .103$, $p = .05$ (*$p = .3$); 95% CI [.000, .405], $d = .321$, $BF_{01} = 1.341$. However, this advantage was not maintained in the Sound Conditions. Familiar distracter stimuli appeared to produce greater disruption than unfamiliar stimuli, but this depended on whether the distracters comprised melody (without lyrics), a combination of lyrics sung to a melody or speech (spoken lyrics).

A 3 (Sound Condition: quiet, unfamiliar, familiar) × 4 (Sound Type: Melody, Familiar Lyrics, Unfamiliar Lyrics, Speech [spoken lyrics]) mixed ANOVA showed a significant main effect of Sound Condition, $F(1.960, 568.49) = 161.226$, $MSE = .230$, $p < .001$, $\eta_p^2 = .357$. There was a significant between-participants main effect of Sound Type, $F(3, 290) = 6.089$, $MSE = 1.335$, $p < .001$, $\eta_p^2 = .059$, and a significant interaction between Sound Condition × Sound Type, $F(5.881, 568.494) = 10.686$, $MSE = .230$, $p < .001$, $\eta_p^2 = .100$. A simple effects analysis (LSD) to

[8]Discrepancies greater than 2 were assessed again for reconsideration.

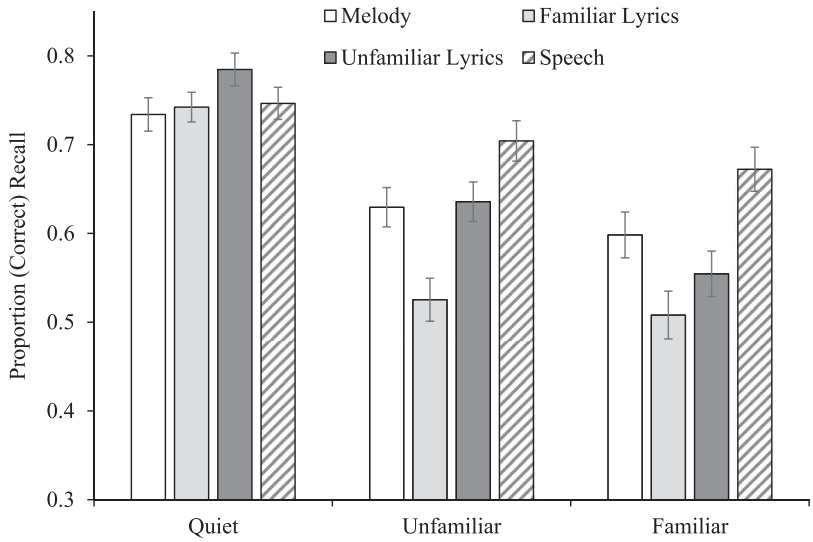


Figure 2. Mean proportion retrieval accuracy for melody humming across the quiet (no sound) and Sound Conditions as a function of Sound Type, deployed in Experiment 1. Error bars represent standard error of the means.

decompose the Sound Condition × Sound Type interaction found that retrieval performance was poorer when accompanied by sound, regardless of Sound Type or whether it comprised a familiar component (e.g. lyrics): Performance in quiet was superior to performance in all other Sound Conditions regardless of Sound Type (all *ps* < .05).

We first compared unfamiliar and familiar Sound Conditions within Sound Type, the only significant difference to arise was for the Unfamiliar Lyrics Group ($MD = -.325$, $SE = .081$, $p < .001$; 95% CI [−.485, −.165], $dz = 0.533$, $BF_{01} = .001$, demonstrating extreme evidence for H1), thereby illustrating a familiarity effect whereby unfamiliar lyrics sung to a familiar melody was more disruptive than unfamiliar lyrics sung to an unfamiliar melody. Next, we compared the impact of unfamiliar, then familiar, stimuli across Sound Type.

Unfamiliar Stimuli. In relation to unfamiliar stimuli, an analysis of Sound Type (e.g. Group) as a function of Sound Condition, found that melody (without lyrics) was more disruptive than speech (spoken lyrics; $MD = -.299$, $SE = .131$, $p = .023$ ($^*p = .069$); 95% CI [−.577, −.041], $d = .392$, $BF_{01} = .570$, indicating anecdotal evidence for H1). Further, sung familiar lyrics was more disruptive than melody (without lyrics; $MD = -.417$, $SE = .129$, $p = .001$ ($^*p = .004$), 95% CI [−.670, −.164], $d = .516$, $BF_{01} = .075$, indicating strong evidence for H1), and sung unfamiliar lyrics ($MD = -.442$, $SE = .129$, $p < .001$, 95% CI [−.695, −.189], $d = .546$, $BF_{01} = .045$, indicating strong evidence for H1) and speech (spoken lyrics; $MD = -.715$, $SE = .129$, $p < .001$, 95% CI [−.968, −.462], $d = .876$, $BF_{01} = .000$, indicating extreme evidence for H1). Finally, sung unfamiliar lyrics was also significantly more disruptive than speech (spoken lyrics; $MD = -.274$, $SE = .131$, $p = .038$ ($^*p = .076$), 95% CI [−.532, −.016], $d = .358$, $BF_{01} = .875$, demonstrating anecdotal evidence for H1).

Familiar stimuli. In relation to familiar stimuli, melody (without lyrics) was more disruptive than speech (spoken lyrics; $MD = -.296$, $SE = .148$, $p = .046$ ($^*p = .138$); 95% CI [−.587, .005], $d = .344$, $BF_{01} = 1.027$, indicating anecdotal evidence for H1). Further, sung familiar lyrics was more disruptive than melody (without lyrics; $MD = -.361$, $SE = .145$, $p = .013$ ($^*p = .052$), 95% CI [−.646, −.076], $d = .394$, $BF_{01} = .506$, indicating anecdotal evidence for H1), and speech (spoken lyrics; $MD = -.657$, $SE = .145$, $p < .001$, 95% CI [−.942, −.372], $d = .730$, $BF_{01} = .001$, demonstrating extreme evidence for H1). Finally, sung unfamiliar lyrics produced greater disruption than speech (spoken lyrics; $MD = -.471$, $SE = .148$, $p = .002$ ($^*p = .01$), 95% CI [−.762, −.180], $d = .550$, $BF_{01} = .050$, indicating strong evidence for H1).

Mean onset time

Mean onset times as a function of Sound Condition and Sound Type can be observed in Figure 3. A 3 (Sound Condition: quiet, unfamiliar, familiar) × 4 (Sound Type: Melody, Familiar Lyrics, Unfamiliar Lyrics, Speech [spoken lyrics]) mixed ANOVA

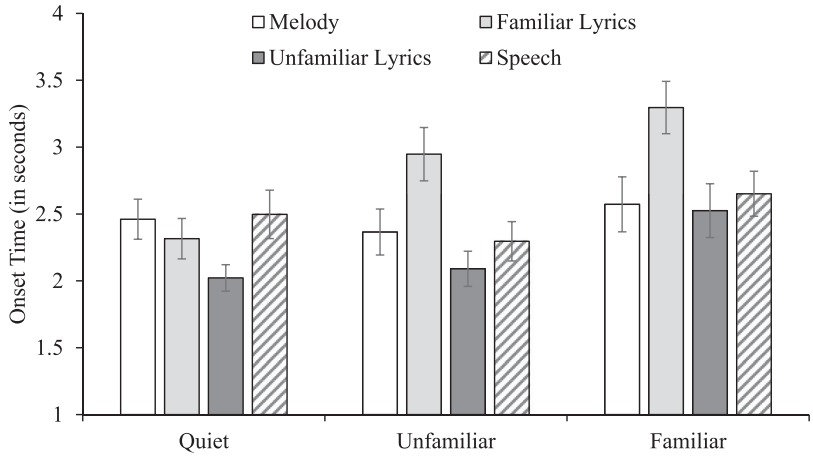


Figure 3. Mean onset time (in seconds) for melody humming retrieval across Sound Condition as a function of Sound Type deployed in Experiment 1. Error bars represent standard errors of the means.

showed a significant main effect of Sound Condition, $F(1.942, 563.235) = 12.587$, $MSE = 1.259$, $p < .001$, $\eta_p^2 = .042$, and Sound Type, $F(3, 290) = 3.981$, $MSE = 3.955$, $p = .008$, $\eta_p^2 = .040$. Furthermore, the Sound Condition × Sound Type interaction was significant, $F(5.827, 563.235) = 3.185$, $MSE = 1.259$, $p = .005$, $\eta_p^2 = .032$.

A simple effects analysis (LSD) was undertaken for the significant Sound Condition × Sound Type interaction. For the sung familiar lyrics condition, longer onset times were revealed for the unfamiliar condition compared to the quiet condition ($MD = -.632$, $SE = .158$, $p < .001$, 95% CI [−.943, −.322], $dz = 0.378$, $BF_{01} = .065$, indicating anecdotal evidence for H1), and for the familiar condition compared to the quiet condition ($MD = -.981$, $SE = .191$, $p < .001$, 95% CI [−1.356, −.605], $dz = 0.601$, $BF_{01} = .000$, demonstrating extreme evidence for H1). There was a tendency for onset times to be longer for the familiar as compared to the unfamiliar condition ($MD = -.349$, $SE = .181$, $p = .055$, 95% CI [−.007, −.705], $dz = 0.200$, $BF_{01} = 2.474$, indicating anecdotal evidence for H0). Further, for the sung unfamiliar lyrics condition, onset times were slower for the familiar condition compared to both the quiet ($MD = -.503$, $SE = .199$, $p = .012$ ($^*p = .036$), 95% CI [−.894, −.112], $dz = 0.277$, $BF_{01} = .782$, demonstrating anecdotal evidence for H1) and unfamiliar condition ($MD = -.435$, $SE = .188$, $p = .022$ ($^*p = .044$), 95% CI [−.805, −.064], $dz = 0.271$, $BF_{01} = .862$, indicating anecdotal evidence for H1).

Unfamiliar melody. Onset times were slower for familiar lyrics sung to an unfamiliar melody than for both unfamiliar melody alone ($MD = -.582$, $SE = .233$, $p = .013$ ($^*p = .052$), 95% CI [−1.041, −.124], $d = .359$, $BF_{01} = .807$, demonstrating anecdotal evidence for H1), and unfamiliar lyrics sung to an unfamiliar melody ($MD = .857$, $SE = .233$, $p < .001$, 95% CI [.399, 1.316], $d = .576$, $BF_{01} = .025$, indicating very strong evidence for H1). Further, familiar lyrics sung to an unfamiliar melody resulted in slower onset times than speech (spoken unfamiliar lyrics; $MD = .652$, $SE = .233$, $p = .005$ ($^*p = .025$), 95% CI [.193, 1.110], $d = .424$, $BF_{01} = .333$, demonstrating moderate evidence for H1).

Familiar sound. Onset times were slower for familiar lyrics sung to a familiar melody than to familiar melody alone ($MD = -.724$, $SE = .271$, $p = .008$ ($^*p = .04$), 95% CI [−1.257, .190], $d = .417$, $BF_{01} = .368$, demonstrating anecdotal evidence for H1), unfamiliar lyrics sung to a familiar melody ($MD = .771$, $SE = .271$, $p = .005$ ($^*p = .03$), 95% CI [.238, 1.305], $d = .449$, $BF_{01} = .229$, indicating moderate evidence for H1), and familiar spoken lyrics ($MD = .645$, $SE = .271$, $p = .018$ ($^*p = .072$), 95% CI [.111, 1.178], $d = .405$, $BF_{01} = .435$, demonstrating anecdotal evidence for H1).

Recognition test

The results of the recognition test, which served as a check for the familiarity manipulation, demonstrated that 79% of familiar melodies were known by participants across Groups. Pairwise comparisons for Sound within Groups showed no significant differences between conditions.

Discussion

The results of Experiment 1 demonstrate for the first time that the production of song, via humming, from long-term memory is disrupted by the mere presence of background sound regardless of the type of sound deployed (melody or speech). This suggests that vocal-motor planning in the production of sequence from long-term memory, like vocal-motor processing in the context of visual-verbal serial recall (e.g. Hughes & Marsh, 2017; Jones et al., 2004), is susceptible to disruption from the mere presence of background sound. Crucially, however, in contrast to studies reporting comparable (Jones & Macken, 1993) or greater (Körner et al., 2017) disruption of visual-verbal serial recall by speech than by non-speech sounds, non-speech sounds (instrumental melodies) were more disruptive of humming performance than were speech sounds (spoken lyrics), regardless of their familiarity. This shows that the similarity between the sequential information provided by the irrelevant material, compared to the target melody (greater for irrelevant melody than irrelevant spoken material), exacerbates disruption. Familiar melody in combination with lyrics (regardless of the familiarity of those lyrics) drove additional disruption of melody retrieval. The fact that unfamiliar lyrics combined with familiar melody (i.e. unfamiliar lyrics sung to a familiar melody) were shown to be more disruptive than unfamiliar lyrics combined with unfamiliar melody (i.e. unfamiliar lyrics sung to an unfamiliar melody), suggests incongruity in familiarity between lyrics and melody may also be important in determining disruption. Finally, a striking finding from the analysis of onset times demonstrated that a combination of familiar melody with familiar lyrics (familiar lyrics sung to a familiar melody) slowed initial production of a target melody to a greater extent than all other conditions, suggesting this condition differentially impairs retrieval of the target melody.

The results of Experiment 1, in part, support the interference-by-process component of the perceptual-motor view of memory (Hughes & Jones, 2005; Jones et al., 2006, 2007). Experiment 1 clearly showed familiarity within a component of song (melody or lyrics) to be a potent distracter for melody retrieval, especially when both the melody and associated lyrics components are familiar. This pattern of results suggests disruption related to increased demands on the vocal-motor system resulting from the concurrent involuntary processing of sound sequences.

At the same time, the pattern of findings is inconsistent with interference-by-content views (Neath, 2000; Salamé & Baddeley, 1982), that assume disruption by irrelevant sound occurs at the item-level. According to such a view, familiar melodies should have been no more distracting than unfamiliar melodies, regardless of the presence of lyrics. Since acoustic variability of all musical items in Experiment 1 was carefully controlled, it is unlikely that any additional disruption produced by familiar melody with lyrics (familiar or unfamiliar) over unfamiliar melody with unfamiliar lyrics is attributable to greater acoustic variation (cf. Schlittmeier et al., 2008) and hence a more pronounced (acoustic) interference-by-process (cf. Jones & Tremblay, 2000). It has been suggested that musicians, as compared to non-musicians, use two working memory (WM) systems (Schulze et al., 2011). These purportedly comprise the phonological loop (e.g. Baddeley, 1986)—for rehearsing verbal information—and a tonal loop—supporting pitch rehearsal. It is difficult to see how appeal to a tonal loop could explain the patterns of irrelevant sound disruption observed in Experiment 1 since the pitch information at the item-level is the unit of currency within the tonal loop and the disruption observed in Experiment 1 is attributed to the involuntary processing of sound sequence and post-categorical factors (e.g. familiarity).

In relation to the independence *vs*. integration debate, the finding that lyrics and melody combined within song, familiar or unfamiliar, in Experiment 1 impaired melody retrieval to a greater degree than spoken-lyrics, familiar or unfamiliar, suggests that song is stored differently to verbal information in WM (Berz, 1995). These results appear to support a degree of independence between lyrics and melody (Besson et al., 1998; Besson & Schön, 2001; Peretz et al., 1994). Further, that spoken-lyrics, familiar or unfamiliar, without melody, were less disruptive than familiar melody and sung-lyrics, also suggests a degree of separation.

Experiment 2

The purpose of Experiment 2 was to shift the focal task process to enable further evaluation of the integration debate and an interference-by-process account (cf. Jones & Tremblay, 2000) of the disruption of song production by different components of song.

Instead of requesting the production of melody via humming, Experiment 2 required participants to vocally produce, via speech, the lyrics component of familiar song in the presence of the same irrelevant auditory conditions as for Experiment 1. Changing the characteristics of the focal task is a device frequently used to enable further assessment of the interference-by-process view (Hughes et al., 2007; Marsh et al., 2018). If the same pattern of auditory distraction is obtained when attempting to speak the lyrics song component as found when trying to produce the melody component of familiar song by humming (Experiment 1), then this would support the notion that melody and lyrics appear to be integrated within song (e.g. Schön et al., 2010). If, however, spoken or sung lyrics impair spoken lyrics performance to a greater degree than melody alone, then some evidence would be obtained for the independence of melody and lyrics processing within song (e.g. Besson et al., 1998; Bonnel et al., 2001).

It should also be considered that the additional demands on semantic processing for lyrics assembly/performance may render the task vulnerable to disruption via the presence of meaning within the irrelevant sound (e.g. Jones et al., 2012; Marsh et al., 2009; Marsh & Jones, 2010). Previous studies have shown that impairment to tasks requiring semantic processing, such as the free or, categorically organised, recall of lists of semantically related words (Marsh et al., 2008, 2009, 2014), or the generation of category exemplars from semantic memory (Jones et al., 2012; Marsh et al., 2017), is prevalent when the irrelevant sound contains semantic information (for a related finding, see Meng et al., 2020). This finding suggests that the principle of interference-by process also extends to automatic (applied to to-be-ignored sound) and voluntarily (applied in the context of the memory task) semantic processes (Marsh & Jones, 2010; Meng et al., 2020). According to the notion that a semantic interference-by-process can arise for lexical semantic retrieval, to-be-ignored spoken lyrics should disrupt spoken retrieval of lyrics from a target song, more than to-be-ignored melody (the opposite pattern identified from Experiment 1). However, some caution might be exercised here, since the extent to which familiar melody alone also governs the implicit retrieval of associated lyrics (Pring & Walker, 1994), and in doing promotes additional disruption via a semantic interference-by-process, remains unclear.

Method

Any modifications/differences from Experiment 1 are detailed below.

Participants

Two hundred and eighty-eight adults aged 18–90 ($M = 46.53$, $SD = 22.41$) took part, and a mixed ANOVA for 144 participants aged over 50 years showed Condition × Cognition interaction non-significant. The interaction between Sound Condition and Sound Type detected in Experiment 1 (with retrieval accuracy as dependent variable) had an effect size of $\eta_p^2 = .100$. An apriori power analysis (using G*Power) of ANOVA with repeated, within-between measures, and number of groups set to 4 and number of measurements set to 3 revealed that a sample size of at least 36 participants is needed to detect an interaction effect of this magnitude.

Design

An identical mixed 3 (Sound Condition) × 4 (Sound Type) design was used: The dependent variable being exact recall accuracy of known song lyrics within each bar of the song. A total of 72 participants took part in each Sound Type Condition, respectively.

Procedure

To ensure comparative assessment, each bar of the song's spoken lyrics matched their hummed melody counterpart and were measured with reference to the number of correct lyrics produced in each bar rather than the correct notes of pitch hummed as for Experiment 1.[9] Musically trained rater comparison revealed Cronbach's $\alpha = .994$. There were no non-musical raters, speech assessment considered to be less subjective than melody. Demographic responses,[10] were comparable to Experiment 1.

[9]See Appendix 2.

[10]Prior musical training and current musical participation were non-significant factors. Western musical culture indicated by 91% of participants, 6.9% indicated joint Asian, and 1% joint African.

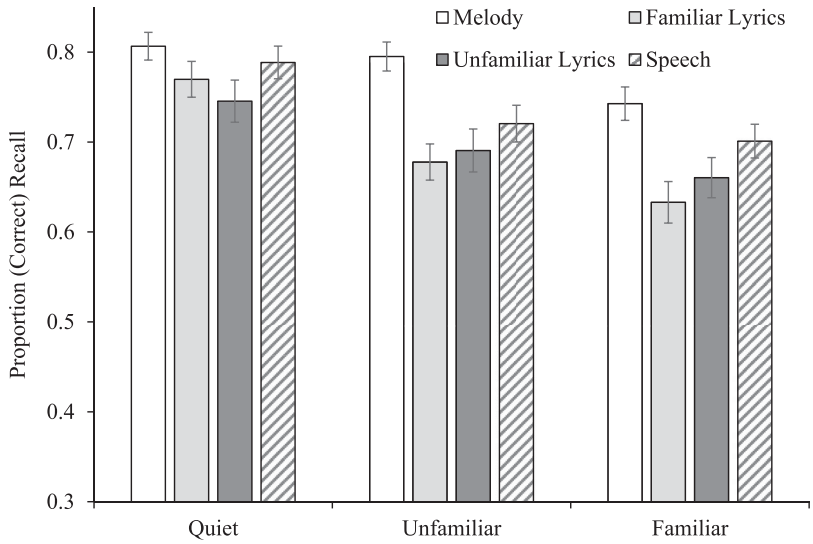


Figure 4. Mean proportion retrieval accuracy for lyrics across the quiet, unfamiliar, and familiar Sound Conditions according to Sound Type in Experiment 2.

Results

Retrieval accuracy

Figure 4 shows the retrieval accuracy of spoken lyrics as a function of Sound Condition and Sound Type in Experiment 2. Retrieval accuracy was higher for all Sound Type Groups in the quiet condition (although significantly lower for the Unfamiliar Lyrics as compared to Melody Group, $p = .027$) identifying the Melody Group as the most able. The disruption of the accuracy of lyric retrieval by to-be-ignored sound appeared to depend on the familiarity of the distracters and whether they comprised melody (without lyrics), lyrics combined with melody or speech (without melody). Accuracy appeared to be impaired to a greater extent in the presence of a familiar, as compared to an unfamiliar, distracter stimulus, and more so when lyrics were familiar (Familiar Lyrics Group). Furthermore, melody (without lyrics) was generally less disruptive than speech (spoken lyrics) but only when the distracter stimuli were unfamiliar.

A 3 (Sound Condition: quiet, familiar, unfamiliar) × 4 (Sound Type: Melody, Familiar Lyrics, Unfamiliar Lyrics, Speech [spoken lyrics]) mixed ANOVA showed a significant main effect of Sound Condition, $F(2, 568) = 54.377$, $MSE = .187$, $p < .001$, $\eta_p^2 = .161$, a between-participants main effect of Sound Type, $F(3, 284) = 5.532$, $MSE = 1.036$, $p = .001$, $\eta_p^2 = .055$, and a significant Sound Condition × Sound Type interaction, $F(6, 568) = 2.371$, $MSE = .187$, $p = .029$, $\eta_p^2 = .024$.

A simple effects analysis (LSD) compared means for Sound Conditions as a function of Sound Type. Performance was better in the quiet condition compared to all other conditions ($ps = <.010$), apart from unfamiliar melody ($MD = .046$, $SE = .073$, $p = .533$; 95% CI [−.009, .190], $dz = 0.090$, $BF_{01} = 8.074$, indicating moderate evidence for H0).

In a first step familiar and unfamiliar Sound Conditions were compared within Sound Type. Performance was poorer in the familiar condition compared to the unfamiliar condition for melody ($MD = .256$, $SE = .074$, $p < .001$; 95% CI [.110, .401], $dz = 0.418$, $BF_{01} = .035$, indicating strong evidence for H1), and familiar lyrics ($MD = .179$, $SE = .069$, $p = .010$; 95% CI [.043, .315], $dz = 0.304$, $BF_{01} = .464$, demonstrating anecdotal evidence for H1). The difference approached significance for unfamiliar lyrics ($MD = .121$, $SE = .069$, $p = .081$; 95% CI [−.015, .257], $dz = 0.173$, $BF_{01} = 3.783$, indicating moderate evidence for H0). Thus, a familiarity effect was observed for unfamiliar compared with familiar melody (without lyrics) and when familiar lyrics were sung to a familiar *versus* an unfamiliar melody. There was a tendency for a familiarity effect to also arise when unfamiliar lyrics were sung to a familiar versus unfamiliar melody. However, no familiarity effect was observed for speech (spoken lyrics). The impact of unfamiliar, then familiar, stimuli was subsequently compared across Sound Type.

Unfamiliar Stimuli. For unfamiliar stimuli, sung familiar lyrics was more disruptive than melody (without lyrics; $MD = -.469$, $SE = .115$, $p < .001$; 95%

CI [−.696, −.243], $d = .759$, $BF_{01} = .001$, indicating extreme evidence for H1), and sung unfamiliar lyrics also more disruptive than melody (without lyrics; $MD = .418$, $SE = .115$, $p < .001$; 95% CI [.192,.644], $d = .604$, $BF_{01} = .018$, demonstrating very strong evidence for H1). Furthermore, speech (spoken lyrics) was also more disruptive than melody (without lyrics; $MD = .299$, $SE = .115$, $p = .010$ ($^*p = .04$); 95% CI [.072, .525], $d = .478$, $BF_{01} = .166$, indicating moderate evidence for H1).

Familiar Stimuli. For familiar stimuli, sung familiar lyrics was more disruptive than melody (without lyrics; $MD = -4.39$, $SE = .118$, $p < .001$; 95% CI [−.670, −.207], $d = .617$, $BF_{01} = .014$, indicating very strong evidence for H1) and speech (spoken lyrics; $MD = .272$, $SE = .118$, $p = .021$ ($^*p = .084$); 95% CI [−5.04, −.041], $d = .381$, $BF_{01} = .656$, indicating anecdotal evidence for H1). Furthermore, sung unfamiliar lyrics was more disruptive than melody (without lyrics; $MD = .329$, $SE = .118$, $p = .005$ ($^*p = .025$); 95% CI [.098, .561], $d = .472$, $BF_{01} = .182$, demonstrating moderate evidence for H1).

Mean onset time

A 3 (Sound Condition: quiet, familiar, unfamiliar) × 4 (Sound Type: Melody, Familiar Lyrics, Unfamiliar Lyrics, Speech [spoken lyrics]) mixed ANOVA showed no significant main effect of Sound Condition, $F(2, 568) = 2.438$, $MSE = 1.226$, $p = .088$, $\eta_p^2 = .009$, no between-participants main effect of Sound Type, $F(3, 284) = .079$, $MSE = 4.439$, $p = .971$, $\eta_p^2 = .001$, and no significant Sound Condition × Sound Type interaction, $F(6, 568) = .645$, $MSE = 1.226$, $p = .694$, $\eta_p^2 = .007$.

Recognition test

Based on the Recognition test, 80% of song lyrics were known by the participants across conditions in Experiment 2. Pairwise comparisons for Sound within Groups showed no significant differences between conditions.

Discussion

The results of Experiment 2 showed target lyrics retrieval—as indexed by the accuracy of lyrics retrieval performance—was adversely affected by all Sound Conditions as compared to quiet (with the exception of unfamiliar melody). Of greater importance, however, was that the results clearly demonstrate that the production of spoken lyrics is more detrimentally affected by sung lyrics than by spoken lyrics or melody thereby replicated the results of Experiment 1. In the absence of lyrics, familiar melody was more potent at impeding spoken lyric production than was unfamiliar melody. Crucially, however, familiar lyrics combined with familiar melody (familiar lyrics sung to an associated familiar melody) was more disruptive than familiar lyrics combined with an unfamiliar melody (familiar lyrics sung to an unfamiliar melody). This suggests that melody familiarity overrode lyrics familiarity if there was an incongruity (familiar lyrics sung to an unfamiliar melody). A striking difference between the patterns of disruption observed between Experiment 1 and Experiment 2 is that unfamiliar speech (spoken lyrics) produced greater disruption of lyric retrieval than unfamiliar melody (without lyrics). The opposite pattern was true in Experiment 1 whereby unfamiliar melody (without lyrics) produced greater disruption than unfamiliar speech (spoken lyrics).

Although some previous studies provide evidence that a common neurological network subserves lexical/phonological and melodic processing (e.g. Schön et al., 2010), the results of Experiment 2 support a degree of processing independence (e.g. Besson et al., 1998; Besson & Schön, 2001; Peretz et al., 1994) at the behavioural level, as here, irrespective of familiarity, sung-lyrics were more disruptive than spoken lyrics of spoken lyrics retrieval. The fact that unfamiliar spoken lyrics produced greater impairment than unfamiliar melody likewise suggests different processes are involved in melody against spoken retrieval, as the opposite pattern of disruption was observed for melody retrieval via humming in Experiment 1. The pattern of disruption is consistent with the notion that the vocal plan for humming, compared to spoken lyrics retrieval, requires melodic, pitch, spectral and temporal information and the processing of these components could be disrupted by similar properties within the to-be-ignored sound (see later).

For Experiments 1 and 2 the passive, interference-by-content view (e.g. Neath, 2000) was pitched against the functional, interference-by-process view (Hughes & Jones, 2005; Jones & Tremblay, 2000). According to the interference-by-content account, whereby disruption occurs at the item-level, familiarity with lyrics or melody should be no more distracting than unfamiliar lyrics/

melodies. Inadequacies of the interference-by-content account are underscored by several findings in the context of Experiment 2. For example, familiar melody produced more disruption than unfamiliar melody despite little overlap in the similarity in content between the task-irrelevant and to-be-recalled material. One possible explanation for the difference in disruption, in line with the interference-by-content view, is that familiar melody automatically activates familiar lyrics and these disrupt retrieval of target lyrics (Pring & Walker, 1994). However, this view does not seem to adequately explain why familiar lyrics combined with familiar melody are more disruptive to lyric retrieval than unfamiliar lyrics combined with familiar melody since in both cases the presence of task-irrelevant lyrics should interfere with the representation and production of task-relevant lyrics. Furthermore, the interference-by-content account does not adequately explain how performance disruption attributable to different properties of task-irrelevant sound are acutely sensitive to the nature of the focal task processing.

The overarching pattern of disruption observed from the Sound Conditions and Sound Type deployed in Experiments 1 and 2 cohere with the suggestion that the presence of to-be-ignored sound containing lyrics impacts upon the efficacy of an articulatory plan responsible for the assembly and production of sequences of lyric patterns. This notion coheres with the perceptual-gestural view whereby the motor planning process operates according to the sequence of items rather than their individuality (e.g. Hughes & Jones, 2005). To-be-ignored sung familiar lyrics in particular, appear to have disrupted the motor-programming necessary to retrieve and perform familiar target melodies and spoken production of lyrics (cf. Leman, 2007; Leman & Maes, 2015). This finding, in part, supports the concept that the motor-planning mechanism underpinning spoken lyric production is more vulnerable to disruption via the presence of sung against spoken lyrics.

The finding that spoken lyric production in Experiment 2 was disrupted more by the presence of speech (spoken lyrics) than melody (without lyrics) when the stimuli were unfamiliar, and that the reverse was true for melody production via humming in Experiment 1, aligns well with the interference-by-process account (Jones & Tremblay, 2000; Marsh et al., 2009). According to this account, the degree and nature of disruption by to-be-ignored sound is jointly dependent on the properties of the sound and the characteristics of the processes underpinning the focal task. The Semantic retrieval element of lyric production, therefore, would render the task more susceptible to disruption via the processing of the semantic, as opposed to the acoustic, properties of to-be-ignored sound (e.g. Jones et al., 2012). Likewise, the melody retrieval element of the humming production task would render the task more susceptible to disruption via the processing of the acoustic features of the to-be-ignored sound (pitch, spectral and temporal information) as opposed to its semantic attributes.

To assess more directly whether effects observed in Experiment 1 (melody retrieval via humming) and Experiment 2 (spoken lyric retrieval) were a consequence of the combination of the characteristics of the focal task and the nature of the to-be-ignored sound, a comparison analysis including data from Experiment 1 and 2 was undertaken.

Comparison analysis

A 3 (Sound Condition: quiet, unfamiliar, familiar) × 2 (Task: humming retrieval *vs.* spoken lyric retrieval) × 4 (Sound Type: Melody, Familiar Lyrics, Unfamiliar Lyrics, Speech [spoken lyrics]) revealed a main within-participant effect of Sound Condition, $F(2, 1148) = 203.863$, $MSE = .207$, $p < .001$, $\eta_p^2 = .262$, and between-participants main effects of Task, $F(1, 574) = 32.803$, $MSE = 1.187$, $p < .001$, $\eta_p^2 = .054$, and Sound Type, $F(3, 574) = 8.269$, $MSE = 1.187$, $p < .001$, $\eta_p^2 = .041$. These results validate the preceding analysis, however, the focus for this analysis was concerned with any interaction between Task and other factors. There was a two-way interaction between Sound Condition and Task, $F(2, 1148) = 20.279$, $MSE = .207$, $p < .001$, $\eta_p^2 = .034$, whereby retrieval of melodies (i.e. via humming) was poorer than retrieval of lyrics for the unfamiliar, $p < .001$; 95% CI [−.523, −.272], $d = .516$, $BF_{01} = .000$ (demonstrating extreme evidence for H1), and familiar, $p < .001$; 95% CI [−.545, −.276], $d = .497$, $BF_{01} = .000$ (indicating extreme evidence for H1), conditions but not for the quiet condition, $p = .052$; 95% CI [−.208, .001], $d = .162$, $BF_{01} = 2.344$ (revealing anecdotal evidence for H0). There was also a significant interaction between Sound Type and Task, $F(3, 574) = 3.300$, $MSE = 1.187$, $p = .020$, $\eta_p^2 = .017$,

whereby performance was poorer with Humming, as compared to the Semantic retrieval (i.e. spoken lyrics output), task for the Melody Group, $p < .001$; 95% CI [−.608, −.205], and the Familiar Lyrics Group, $p < .001$; 95% CI [−.716, −.304], but not the Speech Group, $p = .268$; 95% CI [−.322, .090], nor the Unfamiliar Lyrics Group, $p = .123$; 95% CI [−.368, .044].

Crucially, the three-way interaction between Sound Condition, Task, and Sound Type was significant, $F(6, 1148) = 4.731$, $MSE = .207$, $p < .001$, $\eta_p^2 = .024$. A simple effects analysis (LSD) to decompose the interaction demonstrated significant differences between melody retrieval and lyrics retrieval for unfamiliar melody (without lyrics; $MD = -.663$, $SE = .123$, $p < .001$; 95% CI [−.905, −.420], $d = 1.009$, $BF_{01} = .000$, indicating extreme evidence for H1), and familiar melody (without lyrics; $MD = -.578$, $SE = .134$, $p < .001$; 95% CI [−.840, −.315], $d = .757$, $BF_{01} = .001$, demonstrating extreme evidence for H1). There was a significant difference between melody retrieval and lyrics retrieval for familiar lyrics sung to unfamiliar melody ($MD = -.610$, $SE = .121$, $p < .001$; 95% CI [−.847, −.372], $d = .784$, $BF_{01} = .000$, indicating extreme evidence for H1) and for familiar lyrics sung to familiar melody ($MD = -.500$, $SE = .131$, $p < .001$; 95% CI [−.757, −.242], $d = .572$, $BF_{01} = .027$, demonstrating very strong evidence for H1). Further, there was a significant difference between melody retrieval and lyrics retrieval for unfamiliar lyrics sung to a familiar melody ($MD = -.424$, $SE = .134$, $p = .002$; 95% CI [−.686, −.161], $d = .519$, $BF_{01} = .085$, indicating strong evidence for H1), and there was a tendency towards a difference between melody retrieval and lyrics retrieval when unfamiliar lyrics were sung to an unfamiliar melody ($MD = -.219$, $SE = .123$, $p = .076$; 95% CI [−.462, −.023], $d = .280$, $BF_{01} = .\ 2.027$, demonstrating anecdotal evidence for H1).

Further discussion

The key finding resultant from the comparative analysis was that long-term memory retrieval/performance accuracy for Experiment 1 and Experiment 2 did not show similar patterns of retrieval in the presence of the same Sound Type distracters. The fact that melody retrieval via humming (Experiment 1) was impaired to a greater degree than retrieval of spoken lyrics (Experiment 2), by the same to-be-ignored sound, implies some independence of the processing of melody and lyrics (Besson et al., 1998; Bonnel et al., 2001; Peretz et al., 1994). Of particular interest was that in the context of an unfamiliar stimulus, a dissociation arose between Speech Groups. In Experiment 1, humming a target melody was significantly disrupted more by an unfamilar melody (without lyrics) than by unfamiliar spoken lyrics. In contrast, for Experiment 2, spoken lyrics retrieval performance was significantly disrupted by unfamiliar spoken lyrics as compared with an unfamiliar melody (without lyrics). While onset time was significantly increased by to-be-ignored sung-lyrics for melody retrieval via humming in Experiment 1, no type of sound against quiet produced any differential prolonging of onset time for lyrics retrieval via speaking in Experiment 2.

In sum, the results thus far have been interpreted as illuminating a specific interference-by-process (Jones & Tremblay, 2000) in the context of melody retrieval via humming (Experiment 1) and retrieval of spoken lyrics (Experiment 2). Earlier we described how the vocal-motor system is involved in the planning of melody retrieval and production (cf. Godøy & Leman, 2009; Leman, 2007; Leman & Maes, 2014, 2015) and musical imagery (Halpern, 2001; Smith et al., 1995). For example, Halpern (2001) proposes that the supplementary motor area (SMA) is activated during the mental "replaying of music" and may mediate rehearsal involving motor programmes including imagined humming. Thus, the disruption produced via to-be-ignored sounds could be localised to their degree of competition for the SMA-related processes: Familiar lyrics combined with either familiar or unfamiliar melody, may be the most potent distracter since this overlearned sequence of melody (and lyrics in the case of familiar lyrics) may be a particularly strong competitor to target retrieval requiring melody or spoken lyric production. In the specific instance of exposure to familiar lyrics to an unfamiliar tune it is possible that concurrent parsing of both the auditory unfamiliar melody and the "correct tune" associated with the lyrics could initiate a competing response.

A question remains, however, concerning whether this strong competition for target retrieval imposed by familiar melody combined with familiar/unfamiliar lyrics results in a seizure of the vocal motor system by the to-be-ignored sound regardless of the particular processing required by the focal task. For example, are exactly the same

effects observed for a task that merely requires the visual-verbal serial recall of digits?

As in the case for musical imagery tasks, the SMA has also been implicated in the serial rehearsal of visual-verbal items (Awh et al., 1996; Smith & Jonides, 1998). However, the seriation processes required for the serial recall task may be both quantitatively and qualitatively different from that of melody retrieval via humming and/or retrieval of spoken text. For example, in contrast to retrieval of melody or spoken lyrics, serial recall is characterised by the phenomenological experience of relentless subvocal repetition of the sequence of items. Further, the serial recall task, unlike the melodic and lyrics retrieval task, does not require recall of melodic elements or semantically rich (e.g. lyrics) material. Moreover, familiar melodies are, by definition, learned, habitual sequences while recall of random sequences of letters or digits (typically used as to-be-recalled items in serial recall tasks) are not. The former is typically more disruptive to task performance (Marsh et al., 2013). Thus, the competition for action in serial recall may be driven by different attributes of the to-be-ignored material: its familiarity or the presence of lyrics that interfere with melody and spoken lyrics retrieval may be redundant in disrupting performance on the serial recall task. To address this, a non-music task involving strict serial recall was used in Experiment 3. The deployment of the serial recall task enabled the determination of whether dissociations in susceptibility to task performance disruption from the same irrelevant Sound cCnditions (e.g. familiarity of melody/lyrics) can be modulated by the nature of goal-driven processes adopted (e.g. retrieval of melody/lyrics *vs.* serial rehearsal). Evidence to this effect would gel with predictions of the interference-by-process account (Jones & Tremblay, 2000; Marsh et al., 2009).

Experiment 3

The classical irrelevant sound effect (ISE; Beaman & Jones, 1997) is associated with the serial recall task (Colle & Welsh, 1976; Jones & Macken, 1993; Salamé & Baddeley, 1982). According to the interference-by-process account of the ISE (Jones & Tremblay, 2000), task-irrelevant sounds gain automatic access to a representational system wherein they are organised into coherent streams that flow directly into the articulatory-planning process (Jones et al., 1993). Since interference-by-process results from a conflict of seriation processes, the degree of focal task disruption rests upon the degree of seriation necessary for task performance (Jones & Tremblay, 2000). This view correctly accounts for why little, if any, impact of irrelevant sound is observed on tasks for which serial recall (seriation) is not required (Beaman & Jones, 1997; Klatte et al., 2007). Further, the interference-by-process account is consistent with the finding that the post-categorical content of to-be-ignored sound (e.g. phonology or meaning) is uninfluential in the distraction of serial recall performance (e.g. meaningfulness, Jones et al., 1990). By design, serial recall tasks minimise participants' use of syntactical or lexical semantic processing that might aid ordered recall. Consequentially, participants are considered to co-opt vocal-motor processing and paralinguistic skills to graft transitional probabilities onto items that do not have such properties. However, motor-processing using serial order is open to interference by other serially ordered material that may at some level fit the action-parameters (Hughes & Jones, 2005).

According to the interference-by-process view, the preattentive processing of pre-categorical acoustic changes drives the disruption of serial recall. Therefore, the familiarity of irrelevant auditory distracters should not be influential in the disruption sound conveys to serial recall performance. Of central interest in Experiment 3 is whether the specific patterns of disruption observed from Experiments 1 and 2, including the specific effects of familiarity, are produced due to active engagement in a melody and lyrics retrieval process. The interference-by-process account (e.g. Jones & Tremblay, 2000) assumes that the degree and type of disruption by sound, is dictated jointly by the demands of the focal task, and the characteristics of the to-be-ignored sound. On this account, the same patterns of disruption from the auditory conditions adopted for Experiments 1–2 and observed on melody retrieval (Experiment 1) and lexical (lyrical) retrieval (Experiment 2) should not be observed for the serial recall task that requires vocal-motor processing—subvocalisation—for target items (digits) serial rehearsal, but does not require access to, or production of, song components (e.g. melody/lyrics).

If the results of Experiment 3 are to be consistent with the interference-by-process account (Jones & Tremblay, 2000), then the changing nature of acoustic properties of Sound Condition should be disruptive of serial recall—with irrelevant speech, as compared to tonal melody, being more disruptive (due to its greater acoustic complexity; cf. Tremblay et al., 2000). For example, unlike the pattern of results obtained in Experiment 1, the combination of familiar melody with lyrics should not disrupt the serial recall of digits more than unfamiliar melody with lyrics. This is because the clash of processing in serial recall is due to seriation processes driven by acoustical changes within to-be-ignored sequence and not post-categorical processes related to the familiarity of the task-irrelevant sound.

Method

Participants

Two hundred participants aged 18–88 ($M = 46.55$, $SD = 22.68$) undertook this experiment, with 100 aged over 50 years taking the Addenbrooke's Cognitive Examination Revised (ACR). We used the cut-off point of <82 for inclusion, giving 0.84 sensitivity for dementia, with MMSE >24 showing no cognitive impairment. Although two participants scored below this point removing these participants did not materially affect the pattern of results and conclusions drawn. Demographic outcomes were non-significant.[11] In a study by Sörqvist (2010; Experiment 2) an effect with the size of $d_z = 1.18$ was obtained for the difference in disruptive effects of changing-state tone sequences and steady-state tone sequences on serial recall. This can be taken as indicative of the sample size needed to detect an effect of melodies on serial recall. An apriori power analysis (using G*Power) indicates that a sample size of at least 12 participants is needed to detect an effect of this magnitude.

Design

A 5 (Sound Condition) × 2 (Sound Type) design was adopted. The within-participants variable of Sound Condition was classified into five levels for each of two between participants Sound Type Groups: Group 1, quiet (no distracter), Melody-familiar, Melody-unfamiliar, Unfamiliar sung-lyrics-familiar melody, Unfamiliar sung-lyrics-unfamiliar melody: Group 2, quiet (no distracter), Familiar-sung-lyrics-familiar melody, Familiar sung-lyrics-unfamiliar melody, Speech-familiar, Speech-unfamiliar. Accuracy of digit recall in strict serial order served as the dependent variable.

Materials and apparatus

Digits 1–8 were compiled into 10 series of random presentation for each Sound Condition, controlled by the E-Prime (2) psychology software tool. A confidence continuum allowed for participant self-reported prior knowledge of distracter melodies and lyrics.

Procedure

There were three main modifications to the procedure from Experiments 1–2. First, all participants experienced five, as compared to three, irrelevant Sound Conditions; second, each sound stimulus was reduced from 32 to 10-second (see Klatte et al., 2002) to correspond to focal task duration; and third, irrelevant sound occurred during short-term memoranda presentation, rather than during long-term memory recall/performance.

Participants undertook four practice trials in quiet conditions where they were instructed to memorise, in the order of visual presentation via a computer screen, series of eight single digits drawn from digits 1–8. No digit was repeated within a trial. Prior to each trial a visual orienting + was shown, and instructions to click "Begin Trial" were consistent for each movement through the programme. Digits were presented consecutively at a rate of 1 per second (800-millisecond on, 200-millisecond off) in 72-point equidistant black Monaco font on a white background. From a viewing distance of 45 centimetres each number subtended a vertical visual angle of 1.49° and a horizontal angle of 0.92°. Following a 2-second retention interval at the end of each trial, participants saw a circular array of all 1–8 digits (changed for each trial to eliminate practice order effect) and were required to click on each digit in order of

[11]Western musical culture indicated by 95% of participants, 3% Asian, 1% Oriental, and 1% dual African. Prior musical training/participation were non-significant factors.

presentation. The "**?**" in the centre of the circle to be used if a digit could not be recalled in a specific position. 50 sequences of digits were created by sampling without replacement from the set 1-8. From these, 10 sequences were allocated to each Sound Condition. A confidence continuum was also developed, that allowed for participants' self-reported confidence in their decision, ranging from "Not confident" to "Totally confident". The total task took approximately 45 min. Raw data from each participant was collapsed to establish the mean, accuracy measured using a scale ranging from zero, for no digits, to eight for all digits correctly recalled according to strict serial recall criterion.

Results

Mean serial recall performance as a function of Sound Conditions in Experiment 3 can be observed in Figure 5(Panel A and Panel B).

Following a one-way repeated measures ANOVA, serial-digit recall scores were found to be more accurate in quiet conditions. An unexpected inconsistency in recall score accuracy in the quiet condition between Group 1 and Group 2 created a slight baseline slip (Group 2 had higher performance). However, for both Groups, scores in quiet conditions were significantly higher than those from any of the distracter conditions.

For Group 1 (Melody and Unfamiliar Lyrics), a preliminary one-way repeated measures ANOVA showed a significant main effect of Sound Condition, $F(3.356, 332.210) = 95.185$, $MSE = .012$, $p < .001$, $\eta_p^2 = .490$, whereby all Sound Conditions were significantly different from quiet (all $p < .001$). When comparing familiarity with the presence of unfamiliar lyrics, a 2 ([Melody]Familiarity: Familiar *vs.* Unfamiliar) × 2 (Presence of [Unfamiliar] Lyrics: Present *vs.* Absent) ANOVA revealed a main effect of Familiarity, $F(1, 99) = 4.281$, $MSE = .006$, $p = .041$, $\eta_p^2 = .041$, and a main effect for the Presence of (Unfamiliar) Lyrics, $F(1, 99) = 129.641$, $MSE = .015$, $p < .001$, $\eta_p^2 = .567$. Furthermore, there was a significant (Melody) Familiarity × Presence of (Unfamiliar) Lyrics interaction, $F(1, 99) = 8.333$, $MSE = .006$, $p = .005$, $\eta_p^2 = .078$. The interaction was decomposed with simple effects analysis (LSD). This revealed a significant difference between unfamiliar melody and familiar melody without lyrics ($MD = .038$, $SE = .011$, $p < .001$; 95% CI [.017, .059], Cohen's $dz = 0.354$, $BF_{01} = .037$, indicating strong evidence for H1), but not between unfamiliar melody combined with unfamiliar lyrics and familiar melody combined with unfamiliar lyrics ($MD = -.006$, $SE = .011$, $p = .602$; 95% CI [−.028, .016], Cohen's $dz = 0.052$, $BF_{01} = 11.06$, indicating strong evidence for H0). Thus, a familiar melody only resulted in poorer performance than an unfamiliar melody in the absence of lyrics.

A significant difference was also revealed between familiar melody without lyrics and familiar melody combined with unfamiliar lyrics ($MD = .116$, $SE = .014$, $p < .001$; 95% CI [.087, .145], Cohen's $dz = 0.804$, $BF_{01} = .000$, indicating extreme evidence for H1), and between unfamiliar melody without lyrics and unfamiliar melody combined with unfamiliar lyrics ($MD = .160$, $SE = .014$, $p < .001$; 95% CI [.132, .188], Cohen's $dz = 1.131$, $BF_{01} = .000$, indicating extreme evidence for H1). Thus, the addition of unfamiliar lyrics further impaired performance regardless of the familiarity of the melody with which they were combined.

A preliminary one-way ANOVA for Group 2 (Familiar Lyrics and Speech) also demonstrated a significant main effect of Sound Condition, $F(4, 396) = 112.387$, $MSE = .008$, $p < .001$, $\eta_p^2 = .532$, whereby all Sound Conditions differed significantly from quiet (all $p < .001$). A 2 ([Lyrics] Familiarity: Familiar *vs.* Unfamiliar × Lyrics Type: Sung *vs.* Spoken) revealed no main effect of (Lyrics) Familiarity, $F(1, 99) = .026$, $MSE = .007$, $p = .87$, $\eta_p^2 = .000$. There was a main effect of Lyrics Type, $F(1, 99) = 9.283$, $MSE = .007$, $p = .003$, $\eta_p^2 = .086$, but no interaction between Familiarity and Lyrics Type, $F(1, 99) = .767$, $MSE = .008$, $p = .383$, $\eta_p^2 = .008$. A simple effects analysis revealed that melody combined with (familiar) lyrics was more disruptive than spoken lyrics ($MD = .025$, $SE = .008$, $p = .003$; 95% CI [.009, .041], Cohen's $dz = 0.305$, $BF_{01} = .1564$, indicating moderate evidence for H1).

To investigate any differential impact of melody against spoken lyrics as a function of familiarity, a 2 (Sound Type: Melody[Group 1] *vs.* spoken lyrics [Speech, Group 2]) × 2 (Familiarity: Familiar *vs.* Unfamiliar) mixed ANOVA revealed a within-participant main effect of Familiarity, $F(1, 198) = 7.876$, $MSE = .006$, $p = .006$, $\eta_p^2 = .038$, and a between-participants main effect of Sound Type, $F(1, 198) = 13.872$, $MSE = .055$, $p < .001$, $\eta_p^2 = .065$. Further, there was a marginally significant Familiarity × Sound Type interaction, $F(1, 198) = 3.839$,

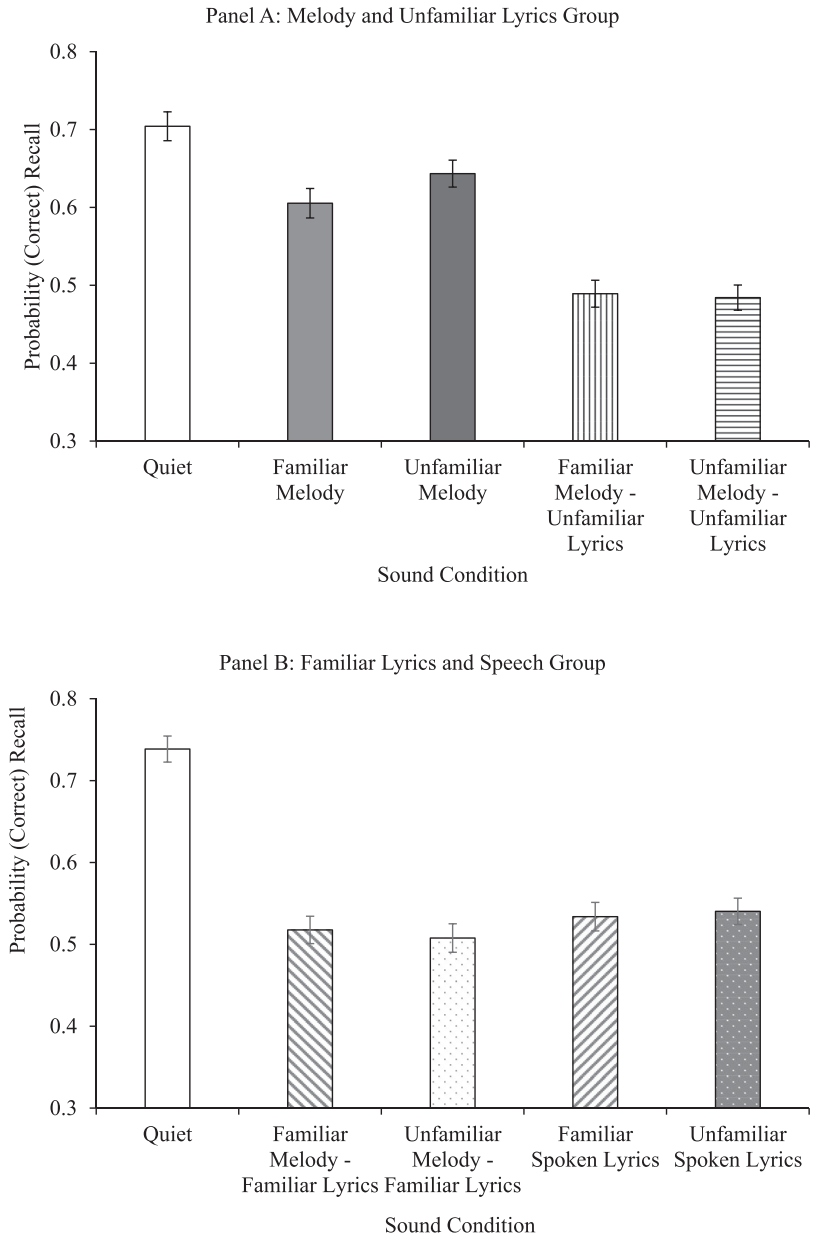


Figure 5. Mean serial recall performance as a function of group (Panel A and Panel B, respectively) and Sound Conditions deployed in Experiment 4. Error bars represent the standard error of the means.

MSE = .006, $p = .051$, $\eta_p^2 = .019$. A simple effects analysis revealed that familiar distracters were more disruptive than unfamiliar when they comprised melody (*MD* = −.038, *SE* = .011, $p < .001$; 95% CI [−.060, .016], Cohen's $d = 0.354$, $BF_{01} = .037$, indicating strong evidence for H1), but not spoken lyrics (*MD* = −.007, *SE* = .011, $p = .55$; 95% CI [−.029, .015], Cohen's $d = 0.057$, $BF_{01} = 10.761$, indicating strong evidence for H0). In addition, Speech (spoken lyrics) was more disruptive than Melody when the distracters were familiar (*MD* = −.072, *SE* = .026, $p = .006$; 95% CI [−.122, −.021], Cohen's $d = 0.393$, $BF_{01} = 0.229$, indicating moderate evidence for H1) and unfamiliar (*MD* = −.103, *SE* = .024, $p < .001$; 95% CI [−.149, −.056], Cohen's $d = 0.614$, $BF_{01} = 0.001$, indicating extreme evidence for H1).

To investigate the potential interplay between melody and lyrics familiarity, a 2 (Melody Familiarity: Familiar *vs.* Unfamiliar) × 2 (Lyrics Familiarity: Familiar [Group2] *vs.* Unfamiliar [Group1]) mixed ANOVA revealed no within-participant main effect of Melody Familiarity, $F(1, 198) = .811$, *MSE* = .007, $p = .369$, $\eta_p^2 = .004$. There was also no between-

Table 2. Mean serial-digit recall scores across Sound Conditions by Sound Type Group.

Group	Condition	Mean	SD	SE
1 / 2	quiet	.702 /.737	.185 / .158	.019 / .016
1	Melody—unfamiliar	.642	.173	.017
1	Melody—familiar	.604	.189	.019
2	Speech—unfamiliar	.540	.161	.016
2	Speech—familiar	.533	.174	.017
2	Familiar Lyrics—familiar melody	.516	.167	.017
2	Familiar Lyrics—unfamiliar melody	.507	.174	.017
1	Unfamiliar Lyrics—familiar melody	.488	.173	.017
1	Unfamiliar Lyrics—unfamiliar melody	.482	.161	.016

Note: Group 1: Melody and Unfamiliar Lyrics; Group 2: Familiar Lyrics and Speech.

participants main effect of Lyrics Familiarity, $F(1, 198) = 1.344$, $MSE = .050$, $p = .248$, $\eta_p^2 = .007$. Further, the Melody Familiarity × Lyrics Familiarity interaction was not significant, $F(1, 198) = .047$, $MSE = .000$, $p = .829$, $\eta_p^2 = .000$.

The main findings were that 1) Spoken lyrics were more disruptive than melody without lyrics regardless of familiarity. 2) Lyrics combined with melody were more disruptive than spoken lyrics and melody without lyrics, regardless of familiarity. 3) Familiar melody without lyrics was more disruptive than unfamiliar melody without lyrics. see Table 2.

The combined Group mean Recognition scores for known melodies was 8.55[12] with 8.48 for known lyrics. A univariate analysis on d'prime scores (sensitivity/discriminability) for melody or spoken lyrics showed no difference between Groups. However, increased confidence was identified, from the c scores, for matched speech (spoken lyrics) compared to matched melody.

Discussion

In line with previous literature using musical excerpts as distracters (Alley & Greene, 2008; Baddeley, 1986; Jones & Macken, 1993; McCorkell, 2012; Perham & Sykora, 2012; Perham & Vizard, 2011; Pring & Walker, 1994; Salamé & Baddeley, 1989; Schlittmeier et al., 2008), serial recall accuracy was highest in quiet conditions as compared to all Sound Conditions, and lowest, irrespective of melody or lyrics familiarity, when combined within song (cf. Iwanaga & Ito, 2002; Nittono, 1997; Salamé & Baddeley, 1989). Consistent with the interference-by-process approach (Jones & Tremblay, 2000), serial recall performance was poorer for spoken lyrics unaccompanied by melody, than for melody alone, suggesting that the acoustic complexity of speech over tonal melody yields a more pronounced interference via competing seriation processes. Note that the opposite pattern is true for the melody retrieval task in Experiment 1, wherein melody was more disruptive than spoken lyrics. Although spoken lyrics were more disruptive than melody in Experiment 2, that required spoken retrieval of lyrics, we assume that the basis of this effect (a semantic interference-by-process) is different from that in the context of serial recall in Experiment 3. Taken together, the opposite findings in relation to task disruption via to-be-ignored melody against spoken lyrics observed in Experiments 1 and 3 coheres with the notion that the deliberate processing of melody in the focal task (Experiment 1), renders the task vulnerable to disruption from the automatic processing of task-irrelevant melody, as the interference-by-process account suggests (Jones & Tremblay, 2000). Also consistent with the interference-by-process account is the failure to replicate the patterns of disruption observed in Experiments 1 and 2. For example, unlike Experiment 1 wherein familiar melody combined with lyrics (either familiar or not) was more disruptive of melody retrieval than unfamiliar melody combined with lyrics, and familiar melody combined with unfamiliar lyrics was more disruptive of melody retrieval than unfamiliar melody combined with unfamiliar lyrics, these sequences were not differentially disruptive of serial recall. Similarly, the lack of any differential disruption according to different sequence types on serial recall differed from the results of Experiment 2 wherein familiar sung lyrics combined with familiar melody was more disruptive of lyrics retrieval than familiar sung lyrics combined with unfamiliar melody. This pattern of results suggests against the notion that familiar melody combined with lyrics usurps control of the vocal-motor system regardless of the focal task process. Rather, it suggests that a more specific competition for action drives the disruptive effect attributable to a combination of familiar melody with lyrics.

[12]Hit rate from 10 items.

The results of Experiment 3 mostly favour the interference-by-process/perceptual-gestural account (Hughes et al., 2009). According to this account, serial recall requires the conversion of visual-verbal sequences into articulatory form, exploiting the speech-planning mechanism, for organisation and maintenance of the required sequence via subvocal rehearsal. Thus, the process required for retention of visual-verbal information across the short-term is independent of melodic/semantic processing. Further, the "perceptual organization" of irrelevant sound via the streaming process operates on pre-categorical acoustic changes. Thus, this interference-by-process mechanism correctly explains why the familiarity of melody/lyrics when combined within task-irrelevant sequences have no disruptive power in the context of serial recall.

The idea that impairment results from requirement to remember and simultaneously ignore similar items that require similar seriation processes—interference-by-content—(Neath, 2000; Salamé & Baddeley, 1982) was not supported from Experiment 3 results. Irrespective of the nature of the sound, serial-digit recall was disrupted compared to recall in quiet conditions.

One unexpected finding from Experiment 3, was that of a distracter familiarity effect whereby familiar as compared to unfamiliar melody was more disruptive of serial recall. This difference occurred even though the changing-state properties (and thus acoustic variability) of both sequence types was equated—thus their disruptive potency should be similar. This distracter familiarity effect is also somewhat puzzling given that the effect of familiarity failed to materialise when melody was combined with lyrics. At first glance it might appear that the presence of to-be-ignored speech (either familiar or unfamiliar) overrides the disruption produced by familiar melody, possibly due to its presence increasing the changing-state of the sequence and thus its power to disrupt serial recall (Jones, 1994). However, that familiar to-be-ignored spoken lyrics failed to disrupt performance relative to unfamiliar to-be-ignored spoken lyrics suggests some limit to the notion that increasing changing-state information eliminates the unique disruption attributable to distracter familiarity. The cognitive underpinnings of the melody familiarity effect observed in Experiment 3, therefore, requires further investigation.

A pivotal question emerging from the results of Experiment 3 is whether the melody familiarity effect observed for serial recall in Experiment 3, can be reconciled with the interference-by-process view. Any explanation of this effect in terms of the interference-by-process account must go beyond simplistic conceptualisation of the changing-state effect: It implies some post-categorical processing (e.g. identification of a melody). Adhering to the framework of interference-by-process, one potential explanation for the melody familiarity effect is that familiar melodies, on account of being overlearned, may strongly activate serial order representations that act as a stronger competing subvocal motor plan (Lima et al., 2016), thereby exacerbating interference-by-(seriation)-process. Further, schema-driven processing (Bregman, 1990) may be invoked by a to-be-ignored familiar melody which gives rise to a competing serial order representation that is stronger than that derived from non-schema driven processing of an unfamiliar to-be-ignored melody: Since the strength of order cues dictates the magnitude of interference-by-process, it is possible that greater disruption of serial digit recall should be observed for sequences of familiar against unfamiliar melody. Yet, the distracter familiarity effect might be better explained by a separate mechanism altogether.

While a debate continues regarding whether the changing-state effect reflects attentional capture (e.g. Körner et al., 2018; Labonté et al., 2021), there is consensus that some forms of auditory distraction reflect such attentional diversion. For example, the duplex-mechanism account argues for the existence of two discrete forms of auditory distraction (Hughes et al., 2007, 2013; Sörqvist, 2010): One is attributable to interference-by-process, the other reflects attentional capture (Hughes et al., 2005; Hughes et al., 2007; but see Bell et al., 2010, 2012). Interference-by-process is manifest through the changing-state effect which, in the context of Experiment 3, accounts for why disruption occurs from all sound sequences, regardless of familiarity, compared to quiet (all sound sequences satisfy the criteria for changing-state). However, in the view of the duplex mechanism account (Hughes, 2014), the distracter familiarity effect (whereby familiar as compared to unfamiliar melody was more disruptive of serial recall) may require appeal to an attentional diversion mechanism. Two forms of attentional capture have been outlined. First, a*specific* attentional capture can occur from a sound that violates expectation (e.g. Parmentier

et al., 2018). This form of attentional capture occurs even when there is no relationship between sound properties and those of task. Second, *specific* attentional capture occurs when the sound's content has the capacity to divert attention and may be independent of the task (Hughes, 2014). For example, increased task errors may be incurred when the background sound's content is responsible for gifting the sound its power to divert attention, irrespective of the processing involved in that task (Hughes et al., 2007; Vachon et al., 2017). For example, by a high valence word (Buchner et al., 2004; Marsh et al., 2018), or by its personal significance (Röer et al., 2013). In the present context, the additional disruption from familiar relative to unfamiliar melody observed in Experiment 3 may cohere with the notion that familiar, as compared with unfamiliar melody, produces specific[13] attentional capture that is independent of the requirements of the task (e.g. Röer et al., 2013; Marsh et al., 2018).

Experiment 4

In Experiment 4, we deploy a task devoid of serial recall, the missing-item task, to assess the mechanism underpinning the greater disruption produced by familiar *versus* unfamiliar melody. The missing-item task has been shown to be sensitive to disruption via properties of sound that induce attentional diversion (e.g. Hughes et al., 2007; Marsh et al., 2018) but is immune to disruption produced by changing sequences of sound (e.g. Beaman & Jones, 1997). The rationale for adopting the missing-item task was that if familiar melody seizes control of, or recruits, motor-planning systems more than unfamiliar melody, then familiar melody should only impair serial digit recall and not the missing-item task, which does not require such sequential motor-planning. Establishing the missing-item task to be invulnerable to the effect of distracter familiarity would lend support to the perceptual-gestural account (albeit requiring an explanation for why enhanced seriation derives from, for example, schema-driven processing of melody), while finding a disruptive effect of melody familiarity would support an attentional diversion account of the distracter familiarity effect in the context of short-term memory.

Method

Only melody distracter material was delivered for this experiment as the familiarity effect identified in Experiment 3 was observed for the Melody Sound Type Group. Any modification to design and procedures from Experiment 3 are detailed below.

Participants

One hundred and six adults aged 18–85 ($M = 43.62$, $SD = 23.04$) participated, 44 of whom were aged over 50 years.[14] For theoretical reasons explicated by the interference-by-process account of auditory distraction, there should be no effect of sound on missing-item task performance, unless the sound sequence comprise deviant (e.g. surprising) sound elements. A power analysis of the sample size needed to detect an effect of sound on the missing-item task is therefore, arguably, better conducted based on data that revealed such an effect and that is theoretically consistent with the interference-by-process account. In a study by Hughes et al. (2007; Experiment 2), an effect of background sound (comprising deviating sound elements) on missing item task performance was reported with a size of $d_z = 1.12$. An apriori power analysis (using G*Power) revealed that a sample size of at least 13 participants is needed to detect this effect.

Design

As a within-participants design, the Melody distraction variable was classified: quiet (no distracter), unfamiliar and familiar. Two blocks of three conditions yielded six orders of presentation, fully counterbalanced. The single item response time was not computed.

Materials and apparatus

In order to yield a richer set of data an additional set of ten familiar melodies (without their associated lyrics), and an additional ten unfamiliar matched melodies created a second block of trials of identical structure to the original distracter block.

[13]As compared with *aspecific* attentional capture: a violation to the expected sound, e.g. "*DDDD**K**DD*".
[14]Scores ranged between 84 and 99 ($M = 94.23$, $SD = 3.26$), MMSE ($M = 28.86$, $SD = 1.19$) positive cognition.

Procedure

Following completion of demographic[15] and consent forms, five practice trials in quiet conditions were undertaken. Participants were visually presented with series of eight single digits, drawn from digits 1–9 (each digit presented once in each trial) and directed to identify the missing digit. Digits were presented consecutively at a rate of 1 per second (800-millisecond on, 200-millisecond off) in 72-point equidistant black Monaco font on a white background. Following a 200-millisecond retention interval at the end of each trial, all nine digits (1–9) were presented on the computer screen with instruction to double click on the digit that was missing from the original trial sequence. There was no time-limit on recall. Ten experimental trials were then delivered for each of six condition blocks (2 each in quiet, familiar melody, unfamiliar melody). Participants were informed they should ignore any sounds being played through headphones; onset of the to-be-ignored sound was simultaneous with to onset of each visual digit pattern presentation. A 30-second rest between experimental conditions was permitted, if needed. Following all experimental trials, participants indicated, from a given list, the strategy they had used to complete the missing item task (based on Morrison et al., 2016).

Results

Overall scores in quiet conditions were the highest (maximum 1). Performance decreased in the unfamiliar melody condition with a further decrement to scores in the familiar melody condition (see Figure 6).

A repeated measures ANOVA showed that overall there was a significant main effect for Sound Condition, $F(2, 210) = 8.544$, $MSE = .013$, $p < .001$, $\eta_p^2 = .075$. Pairwise comparisons demonstrated better performance for quiet compared with familiar melody ($MD = .063$, $SE = .016$, $p < .001$; 95% CI [.030, .095], Cohen's $d = 0.37$, $BF_{01} = 0.016$, indicating very strong evidence in favour of H1), and performance in the unfamiliar melody condition was superior to performance in the familiar melody condition ($MD = .044$, $SE = .016$, $p = .007$ (*$p = .014$); 95% CI [−.013, .076], Cohen's $d = 0.27$, $BF_{01} = 0.327$, indicating moderate evidence for H1), thereby demonstrating a familiarity effect. There was no significant difference between quiet and unfamiliar melody ($MD = .018$, $SE = .014$, $p = .200$; 95% CI [−.010, .047], Cohen's $d = 0.125$, $BF_{01} = 5.752$, indicating moderate evidence for H0) suggesting that the presence of sound *per se* did not produce significant disruption on the missing-item task.

Several participants self-reported using a grouping strategy that possibly indicated that they had undertaken some serial rehearsal (cf. Hughes & Marsh, 2020). Therefore, the means for these two groups self-reported differential use on rehearsal were compared (Rehearsal *versus* non-rehearsal strategy). The ensuing ANOVA revealed no between-participant main effect of Self-Reported Strategy, $F(1, 104) < 0.001$, $MSE = .090$, $p = .999$, $\eta_p^2 < .001$, and no Sound Condition × Strategy interaction, $F(2, 208) = 0.697$, $MSE = .013$, $p = .499$, $\eta_p^2 = .007$. Total Recognition test scores yielded a hit-rate of 7.2[16] and the d'prime measure of recognition and the criterion shift score showed that participants were able to discriminate unfamiliar from familiar melodies, albeit conservatively (see Benjamin & Bawa, 2004).

Interim discussion

The results of Experiment 4 were unequivocal: the missing-item task, widely thought to involve non-seriation processes (e.g. Beaman & Jones, 1997; Buschke & Hinrichs, 1968), was indeed vulnerable to the distracter familiarity effect. Consequently, this vulnerability supports the idea that the mechanism underlying the effect is *specific* attentional capture (Marsh et al., 2018), and not interference-by-process (Jones & Tremblay, 2000). Instead, the results seem to favour an attentional capture account of the familiarity effect. Moreover, further evidence in support of the attentional capture view was gleaned from the fact that additional familiar over unfamiliar melody disruption occurred regardless of whether participants self-reported a seriation or non-seriation-based strategy, as reflected in the absence of a Sound-Condition × Self-Reported Strategy interaction. Thus, the distracter familiarity effect appears task-process insensitive. In contrast, task processing sensitivity appears to be operating for the disruption unfamiliar

[15]100% of participants indicated a "Western" musical culture on their individual response sheet, with dual cultures identified by 4.2%.
[16]Hit rate from 20 items.

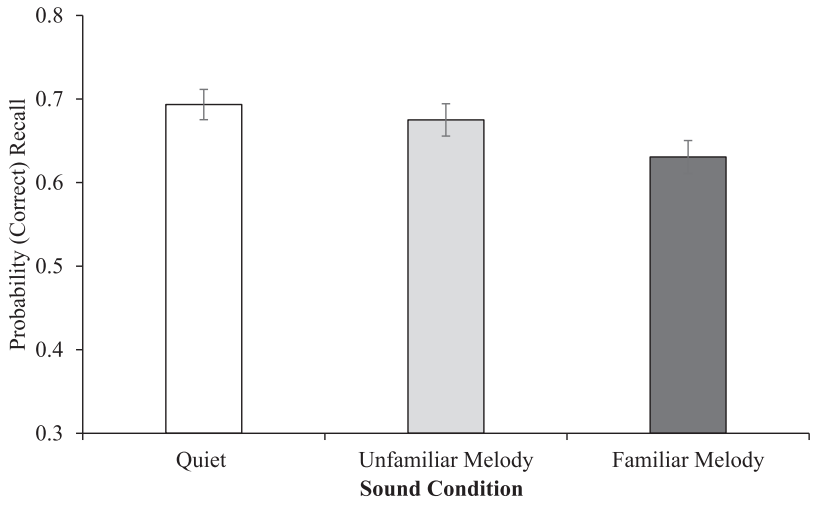


Figure 6. Probability of correct recall as a function of irrelevant Sound Condition for the missing-item task in Experiment 4. Error bars represent standard errors of the means.

melody produces. It appears to be a more pronounced an effect on serial digit recall (Experiment 3) than on the missing-item task (Experiment 4). This was addressed below, with an analysis that directly compares the results of Experiment 4 with those from the Melody conditions of Experiment 3, that required serial recall.

Comparison analysis

A 2 (Sound Conditions: quiet, unfamiliar melody) × 2 (Task: missing-item, serial recall) mixed factor ANOVA showed a significant main effect for Sound Condition, $F(1, 204) = 16.487$, $MSE = 0.010$, $p < .001$, $\eta_p^2 = .075$. There was no between-participants main effect of Task, $F(1, 204) = 0.245$, $MSE = 0.060$, $p = .621$, $\eta_p^2 = .001$. However, the Sound Condition × Task interaction was significant, $F(1, 204) = 4.645$, $MSE = 0.010$, $p = .032$, $\eta_p^2 = .022$. A follow-up simple effects analysis (LSD) confirmed that unfamiliar melody disrupted the serial recall task ($MD = .060$, $SE = .014$, $p < .001$; 95% CI [.033, .087]), Cohen's $d = 0.465$, $BF_{01} = .001$, indicating moderate evidence for H1, but not the missing-item task ($MD = .018$, $SE = .013$, $p = .173$; 95% CI [−.008, .045], Cohen's $d = 0.125$, $BF_{01} = 5.752$, indicating extreme evidence for H0). As expected, the results shown in Figure 7 indicate unfamiliar melody only disrupts performance on the focal task that required seriation (Experiment 3), and so is particularly supportive of the interference-by-process account, wherein it can be assumed that the disruption represents the changing-state effect (Jones & Tremblay, 2000).

Further discussion

The pattern of results across both the serial recall and the missing-item task call into question an interpretation that familiar melodies, being overlearned, activate serial-order representations—such as a competing subvocal motor plan (e.g. Lima et al., 2016)—to a greater degree than unfamiliar melody distracters, empowering them with a superior propensity to clash with subvocal motor output organisation underpinning the serial rehearsal process. On this approach, it would be expected that an interference-by-process, attributable to melody familiarity would be exacerbated on performance of serial, rather than non-serial, short-term memory tasks. At odds with this, the melody familiarity effect was observed on the non-seriation-based missing-item task (Beaman & Jones, 1997; Jones & Macken, 1993). Rather, the familiarity effect specifically gels with an attentional diversion view according to which the content of familiar, as compared to unfamiliar, melody, is responsible for diverting attention away from the focal task (i.e. specific attentional capture; cf. Marsh et al., 2018).

General discussion

The results of the current series can be summarised as follows: As indexed by accuracy of melody retrieval via humming, Experiment 1 demonstrated that the mere presence of background sound disrupted the production of song. Regardless of their familiarity, instrumental melodies were more disruptive than spoken lyrics. Familiar melody in combination

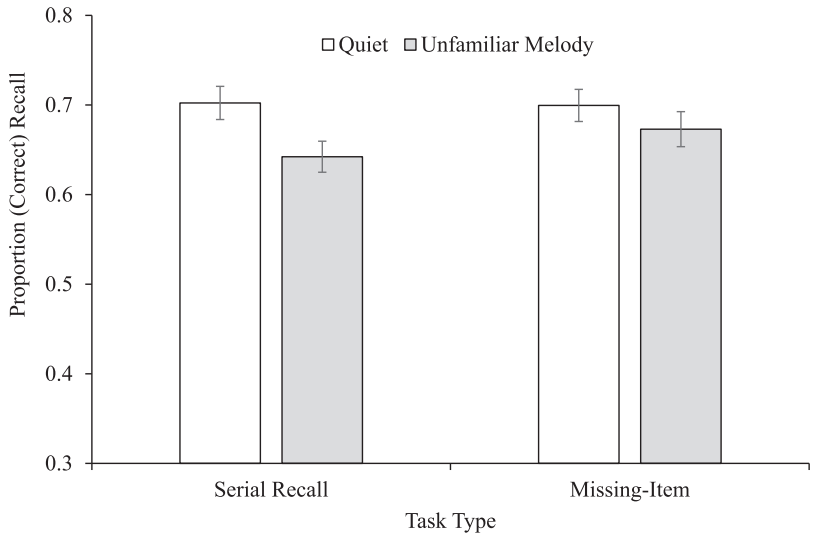


Figure 7. Probability of correct recall as a function of the quiet condition and unfamiliar irrelevant Sound Condition for the serial recall task (Experiment 3) and missing-item task (Experiment 4). Error bars represent standard error of the means.

with familiar or unfamiliar lyrics, produced additional disruption to song retrieval via humming. Furthermore, unfamiliar lyrics combined with familiar melody produced greater disruption than unfamiliar lyrics combined with unfamiliar melody. Onset time for song retrieval was increased when familiar lyrics were combined with melody as compared with all other background Sound Conditions. Similarly, Experiment 2 revealed that the accuracy of the retrieval of target lyrics was disrupted by all background Sound Conditions. Replicating Experiment 1, greater disruption was observed for sung lyrics as compared with spoken lyrics or melody alone. However, familiar melody disrupted spoken lyric production to a greater extent than unfamiliar melody. Familiar lyrics combined with familiar melody disrupted spoken lyrics retrieval more than unfamiliar lyrics combined with an unfamiliar melody. Unfamiliar spoken lyrics produced greater disruption than unfamiliar melody alone. This is at odds with Experiment 1 where the opposite pattern emerged. Experiment 3 revealed that all background sound conditions disrupted the serial recall task. Here, spoken lyrics unaccompanied by melody produced greater disruption than melody alone, an opposite pattern to that found in Experiment 1. Unlike Experiments 1 and 2, no differences emerged between the conditions in which familiar, or unfamiliar lyrics were combined with unfamiliar or familiar melody. However, familiar melody alone produced greater disruption of serial recall than unfamiliar melody alone. This melody familiarity effect was replicated in Experiment 4 wherein the missing-item task that does not necessitate serial order processing was adopted to address whether the manifestation of this melody familiarity depended on the adoption of a seriation strategy.

The results of all the experiments appear to gel well within a process-oriented account of the disruption produced by task-irrelevant material (Hughes & Jones, 2005; Jones & Tremblay, 2000; Marsh et al., 2009; Neumann, 1996). The perceptual-gestural view (Hughes & Marsh, 2017; Jones et al., 2004, 2006) holds that short-term memory is subserved by general purpose perceptual and motor mechanisms wherein inner speech is co-opted to bind together visual-verbal items into motor-plan that enables their sequential reproduction. In the case of the classical ISE, the perceptual processing of serial order in the sound, as part of the perceptual streaming process (Bregman, 1990) conflicts with the vocal-motor processing of serial order in the primary task. In this setting this interference may emerge as the result of the costs of mechanisms that play a role in resolving the competition-for-action promoted by order cues that arise from the deliberate processing of to-be-remembered material and the automatic processing of to-be-ignored material (Hughes & Jones, 2003; Marsh et al., 2009). On this view the semantic properties of background sound, including the familiarity of music and/or lyrics, does not play a role in the disruption (Experiment 3; see also Buchner et al.,

1996; Jones et al., 1990; Marsh et al., 2009). This is because the classic ISE is driven by the information that the background sound conveys in terms of its serial order, not its meaning. Such information is a superficial fit with the action (or vocal-motor process) of serial rehearsal. Due to its greater acoustic complexity than non-speech sounds, background speech conveys more cues as to serial order that enhances its competition for the vocal-motor process underpinning serial recall performance (e.g. Tremblay et al., 2000). The interference-by-process approach also explains why different characteristics of background sound become more disruptive when the focal task processes change (Experiments 1–2; Marsh et al., 2008, 2009; Meng et al., 2020). When, for example, the primary task requires dynamic vocal-motor retrieval processes involving a familiar target melody (Leman, 2007; Leman & Maes, 2014, 2015), irrelevant melodic (e.g. contour) information extracted from background sound produces stronger specific competition for these vocal-motor processes, while spoken lyrics that lack a melodic element produce weaker competition. This finding is reminiscent of previous research demonstrating a modality-specific ISE when comparing tones and speech (LeCompte et al., 1997; Pechmann & Mohr, 1992; Schendel & Palmer, 2007; Williamson et al., 2010). Furthermore, background song wherein one of the two elements (lyrics or melody) is familiar, confers even stronger competition with the vocal-motor melodic retrieval processes because they activate candidate overlearned sequences within long-term semantic memory that compete for retrieval. Slower onset time measures for melody retrieval in the presence of familiar background melody combined with lyrics suggests that vocal-motor planning may be particularly compromised by background familiar song perhaps because it assumes control of the retrieval process and requires removal.

The interference-by-process account also readily explains the shift in the patterns of disruption from background sound that occur when the focal task requires lyrics retrieval (Experiment 2). Here, spoken (unfamiliar) lyrics become more disruptive than (unfamiliar) melody because the primary task involves dynamic retrieval processes focused on lexical-semantic representations. Thus, irrelevant semantic information extracted from the speech produces competition for these processes—a semantic interference-by-process (Marsh et al., 2009). That all Sound Conditions produced disruption relative to quiet suggests that vocal-motor planning is required for the overt production of sequences of spoken lyrics. The greater disruption from song in which one element (lyrics or melody) was familiar compared to when both components were unfamiliar coheres with the notion that this information activates long-term representations of well-known songs that highly specify context-compatible, but ultimately response-inappropriate, information in the context of the lyric retrieval task. The greater disruptive effect of familiar against unfamiliar melody (without lyrics) within Experiment 2, suggests that familiar melody may activate competitor song to a greater extent when the focal task requires lyrics retrieval relative to melody retrieval (Experiment 1). Perhaps this is because melody retrieval via humming requires melodic, pitch, spectral and temporal cues, that are disrupted to a greater extent by cues from any (unfamiliar or familiar) background melody.

One finding that is potentially problematic for the process-oriented approach is that familiar melody (without lyrics) produced greater disruption than unfamiliar melody (without lyrics) in the context of visual-verbal serial recall (Experiment 3). Although this pattern was observed for lyric retrieval in Experiment 2, it may have a different basis in the context of visual-verbal short-term memory. To explore this, in Experiment 4 the impact of familiar versus unfamiliar melody was explored in the context of the missing-item task that arguably does not involve, or at least does not necessitate (Hughes & Marsh, 2020; Morrison et al., 2016) memory for serial order and hence vocal-motor planning. That the missing-item task was shown to be as vulnerable to disruption produced by background familiar against unfamiliar melody suggests that the effect in the context of visual-verbal short-term memory, is underpinned by a mechanism other than interference-by-process, namely specific attentional capture (Marsh et al., 2018). On this view, the content of the background sound has the capability to divert attention away from the focal task, irrespective of the processes underpinning that task (Hughes et al., 2007; Vachon et al., 2017). In the context of visual-verbal serial recall, the changing-state effect and attentional capture can be additive (Hughes et al., 2005, 2007; Marsh et al., 2018). Thus, familiar, and unfamiliar melodies both produce a changing-state effect on serial recall but the additional disruption produced by familiar

melody can be attributed to specific attentional capture (for a similar logic, see Hughes & Marsh, 2019). Specific attentional capture can be due to the content being of personal significance (Röer et al., 2013), or intrigue (Hughes & Marsh, 2020) to the participant, or because it possesses emotional valence for the listener (Buchner et al., 2004; Marsh et al., 2018).

Taken together, however, the results of Experiments 1–4 can be interpreted within a process-oriented approach compared with attentional-resource-based accounts of the disruption produced by background sound (Bell et al., 2019; Cowan, 1995; Lange, 2005; Neath, 2000). For example, on the attentional capture approach (Bell et al., 2019; Cowan, 1995; Lange, 2005), the classical ISE is explained due to acoustic changes-in-state from one item to the next in an irrelevant sequence causing an orienting response towards the sound and thus away from the focal task. On this account it must be assumed that the same patterns of disruption from background sound should be observed regardless of the nature of the focal task-processing. Evidence demonstrating that attentional capture effects are observed on the missing-item task (Beaman & Jones, 1997; Buschke, 1963) that does not engage, or necessitate, a seriation process, while a changing-state effect is not (Hughes et al., 2007; see also Hughes & Marsh, 2020; Marsh et al., 2018, 2023) suggests that attentional capture and changing-state effects are functionally distinct. Similarly, Experiments 1–4 clearly show that the nature of auditory distraction is crucially dependent on the prevailing mental activity further undermines the attentional capture approach while at the same time supporting an axiomatic tenet of the interference-by-process approach (Jones & Tremblay, 2000; Marsh et al., 2008, 2009). Indeed, task-sensitivity to auditory distraction suggests that process, not content, dictates the magnitude and character of disruption and aligns with a functionalist approach to memory in which task-goals and the retrieval environment (instructions, cues, task demands) are central to remembering and forgetting (cf. Toth & Hunt, 1999).

The current study also sought to shed light on the integration/independence debate concerning melody and lyrics within song. If the melody and lexical-semantic (e.g. lyric) retrieval is undertaken within the same processing system, or via shared processes, then comparable patterns of disruption should have been observed between Experiments 1 and 2. However, the retrieval accuracy for melody (Experiment 1) and spoken lyrics (Experiment 2) was influenced differently by the same background sounds, which supports the notion of independence (Besson et al., 1998; Besson & Schön, 2001; Bonnel et al., 2001; Peretz et al., 2009). Greater evidence of independence was provided by the finding that familiar melody (without lyrics) was not significantly more disruptive than unfamiliar melody (without lyrics) for melody retrieval in Experiment 1 but was for lyrics retrieval in Experiment 2. Further, that familiar melody with associated lyrics was indeed more disruptive than unfamiliar melody, and unfamiliar melody with unfamiliar lyrics (Experiment 1), suggests that associated lyrics were not imagined or automatically activated by the presence of familiar melody per se (i.e. without lyrics; see Bailes et al., 2012; Pring & Walker, 1994) as an integration account might predict.

Arguably, several findings reported in Experiments 1 and 2 might be understood within the modular model of music processing (Peretz & Coltheart, 2003). On this account, the musical lexicon contains all the representations of musical phrases that an individual has been exposed to during their lifetime. Recognition of an incoming familiar tune requires selection of that tune from others within the musical lexicon. Output from the musical lexicon, for example, when one is to produce a song requires pairing of the song with lyrics that are stored in the phonological lexicon. Song and lyrics are then integrated and planned for vocal production. It is possible that, consistent with the perceptual-gestural approach (Hughes & Marsh, 2017; Jones et al., 2004, 2006), this planning draws upon a dynamic interplay between music perception, motor-planning and action-planning and reflects embodied cognition (Leman, 2007; Leman & Maes, 2014). On the modular model of music processing, melody production (e.g. via humming, Experiment 1) can occur independently of the phonological lexicon but requires pitch and temporal organisational processes. Activation of the same pitch and temporal processes by a background melody would thus produce difficulty in accessing the musical lexicon for familiar melody as well as its vocal production via humming (Experiment 1). However, spoken lyrics, which primarily activate the phonological lexicon will consequently produce less disruption. Conversely, spoken lyric production makes greater demands on accessing

and output from the phonological lexicon and thus activation of other entries within the phonological lexicon by unfamiliar or familiar spoken lyrics produces greater disruption than melody.

According to the modular model of music processing (Peretz & Coltheart, 2003) bidirectional links exist between the musical lexicon and the phonological lexicon which means that when a familiar background melody activates its representation within the musical lexicon it can also activate its associated lyrics within the phonological lexicon. Similarly, spoken lyrics that activate the phonological lexicon can activate the associated melody within the musical lexicon. That familiar melody against unfamiliar melody produces greater disruption to spoken lyric retrieval (Experiment 2) than melodic retrieval (Experiment 1) suggests that the link between the musical lexicon to the phonological lexicon is stronger than the other way around: That is, familiar melody activates representations of associated lyrics that competes with the vocal planning of spoken lyrics, while spoken lyrics do not (at least strongly) activate melody that can interfere with the vocal planning of melody production via humming. Unfamiliar melody accompanied with unfamiliar lyrics produces less disruption that the other combination of melody and lyrics because both unfamiliar melody and unfamiliar lyrics fail to strongly activate representations of (familiar songs) within the musical and phonological lexicons that could compete with vocal planning. Providing at least one component of the song is familiar, competing representations within the musical lexicon and phonological lexicon, either direct, or through the bilateral link between musical lexicon and phonological lexicon, lead to that representation strongly competing for the vocal-planning underpinning melodic and spoken lyric, retrieval. That is, selection of a familiar target melody or familiar target lyrics is influenced by the activation of competitors within the musical and phonological lexicon. This can have a bearing on the actual selection of target information (melody or lyrics) to populate the motor-planning stage for the target sequence (cf. Peretz et al., 2009; Peretz & Zatorre, 2005). Dynamic selective attentional processes such as inhibition may facilitate target selection by resolving this competition or by removing the candidate, but response inappropriate, melody, or lyrics, from the planning process in the event that they assume the control of action (cf. Neumann, 1987).

The results in relation to onset-time (Experiment 1) may also be explicable within the modular model of music processing (Peretz & Coltheart, 2003). Here, the presentation of a familiar melody with familiar lyrics would result in faster recognition and activation of the competitor within the musical lexicon and thus more completely influence the vocal-planning process, perhaps through capturing the articulators. The finding that familiar lyrics combined with familiar melody delayed onset time for melody retrieval relative to the other background conditions in Experiment 1 is consistent with this explanation: a longer time would be required to resolve the competition and recover the target melody in this condition. That the effect on onset time was not observed for lyrics retrieval in Experiment 2 might be attributable to two explanations. First, when lyrics are decoupled from melody—as in spoken lyric retrieval—there may be less demand on the vocal planning mechanism to produce the output. Second, melody retrieval may be entwined to a greater degree with associated memories than spoken lyric retrieval and as such the task parameters yield greater opportunity generally for background material to compete for, and therefore disrupt, the vocal-planning mechanism responsible for target melody production.

Limitations

Great care was taken to match familiar and unfamiliar melodies and lyrics within the background Sound Conditions used within this experimental series. For this reason, however, it is possible that some unfamiliar background sequences may have been perceived as familiar due to melody priming. To avoid this potential problem in future work, unfamiliar folk song melodies might be adopted as unfamiliar melodies (e.g. Dowling et al., 1995). The results of Experiments 1–4, however, demonstrate that participants were able to discriminate familiar from unfamiliar melodies well enough for them to be differentially disruptive. Indeed, this was also supported by supplementary data from recognition tests. Nevertheless, the use of melodies that are not derived from familiar melodies may produce purer effects. It is also possible that the effects of familiarity might have been diluted somewhat through semantic priming. Some of the songs used may be semantically and associatively related to one another through, for example, theme (e.g.

Christmas, or holiday songs) or learning episode (e.g. during pre-school).

A key question for future research would be to establish whether the production of a target song (melody or lyrics, e.g. "Jingle Bells") is more disruptive by the concurrent presentation of a semantically (and contextually) associated song (e.g. "Rudolph the Red Nose Reindeer") compared to a semantically (and contextually) non-associated song (e.g. "Row, row, row your Boat"). On the modular model of music processing (Peretz & Coltheart, 2003), such categorical priming between musical lexicon representations should exacerbate competition for the vocal-motor planning process.

In Experiment 2 we requested spoken retrieval of lyrics. This task, however, may have involved inhibition of the tendency to sing lyrics accompanied to their melody. An outstanding question, therefore, is whether requesting singing as the vocal-motor output would change the patterns of disruption observed from unfamiliar and familiar melody and lyrics. Singing, as compared with humming (Experiment 1), requires a greater level of vocal co-ordination, due to the simultaneous activation of melody and lyrics. The production of melody and lyrics through singing also offers an advantage of undertaking scoring of these two facets (accuracy of melody and lyric retrieval) on the same output as compared to different outputs in Experiments 1 and 2. Furthermore, it would be interesting to discover if patterns of results from speaking/singing in the presence of different properties of background sound, results in patterns similar to the clinical dissociations identified from neurological studies wherein patients with left frontal lesions can sing, but not speak, words (e.g. Brust, 2003).

It should also be noted that some of our conclusions should be taken tentatively. When adjusting for multiple tests using the Holm–Bonferroni sequential method (Holm, 1979) to deal with family-wise error rates, some of the comparisons in Experiments 1 and 2 approached significance rather than achieved significance. These are indicated in the results sections.

Conclusion

The findings reported here show that retrieval of melody (Experiment 1) and spoken lyrics (Experiment 2) is disrupted by background sound in ways coherent with the dynamic interference-by-process view of the interference-forgetting relationship (Jones & Tremblay, 2000; Marsh et al., 2009). Disruption to the vocal-motor planning process required to retrieve melody or lyrics is produced due to a competition for retrieval that differs in terms of specificity with the requirements of the focal task. As such the explanatory compass of the interference-by-process construct, successful in explaining auditory distraction within short-term serial recall (Jones & Tremblay, 2000), semantic organisation (Marsh et al., 2009), creativity (Marsh et al., 2021), reading (e.g. Meng et al., 2020; Vasilev et al., 2020) and writing (Sörqvist et al., 2012) can be successfully extended to the retrieval of song.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

John E. Marsh's contribution to this article was supported by a grant received from the Bial Foundation [grant number 201/20] and the German Academic Exchange Service (Deutscher Akademischer Austauschdienst [DAAD, grant number 91875827]).

Data availability statement

The data that support the findings of this study are available from the corresponding author, JEM, upon reasonable request.

ORCID

John E. Marsh http://orcid.org/0000-0002-9494-1287

References

Alley, T. R., & Greene, M. E. (2008). The relative and perceived impact of irrelevant speech, vocal music and non-vocal music on working memory. *Current Psychology*, *27*(4), 277–289. https://doi.org/10.1007/s12144-008-9040-z

Anaya, E. M., Pisoni, D. B., & Kronenberger, D. G. (2017). Visual-spatial sequence learning and memory in trained musicians. *Psychology & Music*, *45*(1), 5–21. https://doi.org/10.1177/0305735616638942

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from PET. *Psychological Science*, *7*(1), 25–31. https://doi.org/10.1111/j.1467-9280.1996.tb00662.x

Baddeley, A. D. (1986). *Working memory*. Clarendon Press.

Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford University Press.

Bailes, F., Bishop, L., Stevens, C. J., & Dean, R. (2012). Mental imagery for musical changes in loudness. *Frontiers in Psychology*, *3*, 525. https://doi.org/10.3389/fpsyg.2012.00525

Beaman, C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 459–471. https://doi.org/10.1037/0278-7393.23.2.459

Bell, R., Dentale, S., Buchner, A., & Mayr, S. (2010). ERP correlates of the irrelevant sound effect. *Psychophysiology*, *47*(6), 1182–1191. https://doi.org/10.1111/j.1469-8986.2010.01029.x

Bell, R., Röer, J. P., Dentale, S., & Buchner, A. (2012). Habituation of the irrelevant sound effect: Evidence for an attentional theory of short-term memory disruption. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1542–1557. https://doi.org/10.1037/a0028459

Bell, R., Röer, J. P., Lang, A.-G., & Buchner, A. (2019). Distraction by steady-state sounds: Evidence for a graded attentional model of auditory distraction. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(4), 500–512. https://doi.org/10.1037/xhp0000623

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language*, *51*(2), 159–172. https://doi.org/10.1016/j.jml.2004.04.001

Berz, W. (1995). Working memory in music: A theoretical model. *Music Perception*, *12*(3), 353–364. https://doi.org/10.2307/40286188

Besson, M., Faïta, F., Perez, I., Bonnel, A., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, *9*(6), 494–498. https://doi.org/10.1111/1467-9280.00091

Besson, M., & Schön, D. (2001). Comparison between language and music. *Annals of the New York Academy of Sciences*, *930*(1), 232–258. https://doi.org/10.1111/j.1749-6632.2001.tb05736.x

Bonnel, A. M., Faïta, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception & Psychophysics*, *63*(7), 1201–1213. https://doi.org/10.3758/BF03194534

Bregman, A. S. (1990). Auditory scene analysis: Hearing in complex environments. In S. McAdams & E. Bigand (Eds.), *Thinking in sound* (pp. 10–36). Oxford University Press.

Brust, J. C. M. (2003). The cognitive neuroscience of music. In I. Peretz & R. Zatorre (Eds.), *Music and neurologist: A historical perspective* (pp. 181–191). Oxford University Press.

Buchner, A., Irmen, I., & Erdfelder, E. (1996). On the irrelevance of semantic information for the irrelevant speech effect. *The Quarterly Journal of Experimental Psychology Section A*, *49*(3), 765–779. https://doi.org/10.1080/713755633

Buchner, A., Rothermund, K., Wentura, D., & Mehl, B. (2004). Valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory & Cognition*, *32*(5), 722–731. https://doi.org/10.3758/BF03195862

Buschke, H. (1963). Relative retention in immediate memory determined by the missing scan method. *Nature*, *200*(4911), 1129–1130. https://doi.org/10.1038/2001129b0

Buschke, H., & Hinrichs, J. V. (1968). Relative vulnerability of item information in short-term storage for the missing scan. *Journal of Verbal Learning and Verbal Behaviour*, *7*(6), 1043–1048. https://doi.org/10.1016/S0022-5371968)80065-9

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 17–31. https://doi.org/10.1016/S0022-5371(76)90003-7

Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology*, *55*(1), 75–84. https://doi.org/10.1111/j.2044-8295.1964.tb00899.x

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford University Press.

Crowder, R. G., Serafine, M. L., & Repp, B. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory & Cognition*, *18*(5), 469–476. https://doi.org/10.3758/BF03198480

Dowling, W. J., Kwak, S.-Y., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics*, *57*(2), 136–149. https://doi.org/10.3758/BF03206500

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fiveash, A., McArthur, G., & Thompson, W. F. (2018). Syntactic and non-syntactic sources of interference by music on language processing. *Scientific Reports*, *8*(1), 1–15, [17918]. https://doi.org/10.1038/s41598-018-36076-x

Furnham, A., & Strbac, L. (2002). Music is as distracting as noise: The differential distraction of background music and noise on the cognitive test performance of introverts and extroverts. *Ergonomics*, *45*(3), 203–217. https://doi.org/10.1080/00140130210121932

Godøy, R. I., & Leman, M. (2009). *Musical gestures: Sound, movement, and meaning*. Routledge.

Gröussard, M., Viader, F., Hubert, V., Landeau, B., Abbas, A., Desgranges, B., Eustache, E., & Platel, H. (2010). Musical and verbal semantic memory: Two distinct neural networks? *NeuroImage*, *49*(3), 2764–2773. https://doi.org/10.1016/j.neuroimage.2009.10.039.inserm-00538395

Halász, V., & Cunnington, R. (2012). Unconscious effects of action on perception. *Brain Sciences*, *2*(2), 130–146. https://doi.org/10.3390/brainsci2020130

Halpern, A. (2001). Cerebral substrates of musical imagery. In *Faculty contribution to books* (pp. 127). https://digitalcommons.bucknell.edu/fac_books/127.

Hamzelou, J. (2010, March 9). Music and lyrics: How the brain splits songs. *New Scientist*.

Hébert, S., & Peretz, I. (2001). Are text and tune of familiar songs separable by brain damage? *Brain and Cognition*, *46*(1–2), 169–175. https://doi.org/10.1016/S0278-2626(01)80058-0

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70. https://doi.org/10.2307/2532027

Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCH Journal*, *3*(1), 30–41. https://doi.org/10.1002/pchj.44

Hughes, R. W., Chamberland, C., Tremblay, S., & Jones, D. M. (2016). Perceptual-motor determinants of auditory-verbal serial short-term memory. *Journal of Memory and Language*, *90*(Supplement C), 126–146. https://doi.org/10.1016/j.jml.2016.04.006

Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Vachon, F., & Jones, D. M. (2013). Cognitive control of auditory distraction: Impact of task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 539–553. https://doi.org/10.1037/a0029064

Hughes, R. W., & Jones, D. M. (2003). A negative order-repetition priming effect: Inhibition of order in unattended auditory sequences? *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 199–218. https://doi.org/10.1037/0096-153.29.1.199

Hughes, R. W., & Jones, D. M. (2005). The impact of order incongruence between a task-irrelevant auditory sequence and a task-relevant visual sequence. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 316–327. https://doi.org/10.1037/0096-1523.31.2.316

Hughes, R. W., & Marsh, J. E. (2017). The functional determinants of short-term memory: Evidence from perceptual-motor interference in verbal serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 537–551. https://doi.org/10.1037/xlm0000325

Hughes, R. W., & Marsh, J. E. (2020). When is forewarned forearmed? Predicting auditory distraction in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0000736

Hughes, R. W., Marsh, J. E., & Jones, D. M. (2009). Perceptual-gestural (mis)mapping in serial short-term memory: The impact of talker variability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1411–1425. https://doi.org/10.1037/a0017008

Hughes, R. W., Vachon, F., & Jones, D. M. (2005). Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 736–749. https://doi.org/10.1037/0278-7393.31.4.736

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1050–1061. https://doi.org/10.1037/0278-7393.33.6.1050

Iwanaga, M., & Ito, T. (2002). Disturbance effect of music on processing of verbal and special memories. *Perceptual Motor Skills*, *94*(3_suppl), 1251–1258. https://doi.org/10.2466/pms.2002.94.3c.1251

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Clarendon Press.

Johnston, W. A., & Strayer, D. L. (2001). A dynamic, evolutionary perspective on attentional capture. In C. Folk & B. Gibson (Eds.), *Attraction, distraction, and action: Multiple perspectives on attentional capture* (pp. 375–397). Elsevier Science.

Jones, D. M. (1993). Objects, streams and threads of auditory attention. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control: A tribute to Donald Broadbent* (pp. 87–104). Clarendon Press.

Jones, D. M. (1994). Disruption of memory for lip-read lists by irrelevant speech: Further support for the changing-state hypothesis. *The Quarterly Journal of Experimental Psychology, Section A*, *47*(1), 143–160. https://doi.org/10.1348/000712699161314

Jones, D. M., Beaman, C. P., & Macken, W. J. (1996). The object-oriented episodic record model. In S. Gathercole (Ed.), *Models of short-term memory* (pp. 209–238). Lawrence Erlbaum Associates.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory and Language*, *54*(2), 265–281. https://doi.org/10.1016/j.jml.2005.10.006

Jones, D. M., Hughes, R. W., & Macken, W. J. (2007). The phonological store abandoned. *Quarterly Journal of Experimental Psychology*, *60*(4), 497–504. https://doi.org/10.1080/17470210601147598

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 369–381. https://doi.org/10.1037/0278-7393.19.2.369

Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition*, *21*(3), 318–328. https://doi.org/10.3758/BF03208264

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 656–674. https://doi.org/10.1037/0278-7393.30.3.656

Jones, D. M., Marsh, J. E., & Hughes, R. W. (2012). Retrieval from memory: Vulnerable or inviolable? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 905–922. https://doi.org/10.1037/a0026781

Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, *4*(2), 89–108. https://doi.org/10.1002/acp.2350040203

Jones, D. M., & Tremblay, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin & Review*, *7*(3), 550–558. https://doi.org/10.3758/BF03214370

Klatte, M., Lee, N., & Hellbrück, J. (2002). Effects of irrelevant speech and articulatory suppression on serial recall of heard and read materials. *Psychologische Beiträgc*, *44*, 166–186.

Klatte, M., Meis, M., Sukowski, H., & Schick, A. (2007). Effects of irrelevant speech and traffic noise on speech perception and cognitive performance in elementary school children. *Noise and Health*, *9*(36), 64–74. https://doi.org/10.4103/1463-1741.36982

Körner, U., Röer, J. P., Buchner, A., & Bell, R. (2017). Working memory capacity is equally unrelated to auditory distraction by changing-state and deviant sounds. *Journal of Memory and Language*, *96*, 122–137. https://doi.org/10.1016/j.jml.2017.05.005

Körner, U., Röer, J. P., Buchner, A., & Bell, R. (2018). Time of presentation affects auditory distraction: Changing-state and deviant sounds disrupt similar working memory processes. *Quarterly Journal of Experimental Psychology*, *72*(3), 457–471. https://doi.org/10.1177/1747021818758239

Labonté, K., Marsh, J. E., & Vachon, F. (2021). Distraction by auditory categorical deviations is unrelated to working memory capacity: Further evidence of a distinction between acoustic and categorical deviation effects. *Auditory Perception & Cognition*, *4*(3–4), 139–164. https://doi.org/10.1080/25742442.2022.2033109

Lange, E. B. (2005). Disruption of attention by irrelevant stimuli in serial recall. *Journal of Memory and Language*, *53*(4), 513–531. https://doi.org/10.1016/j.jml.2005.07.002

LeCompte, D. C. (1996). Irrelevant speech, serial rehearsal and temporal distinctiveness: A new approach to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1154–1165. https://doi.org/10.1037/0278-7393.22.5.1154

LeCompte, D. C., Neely, C. B., & Wilson, J. R. (1997). Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 472–483. https://doi.org/10.1037/0278-7393.23.2.472

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Leman, M. (2007). *Embodied music cognition and mediation technology*. MIT Press.

Leman, M., & Maes, P. J. (2014). Music perception and embodied music cognition. In L. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 81–89). Routledge.

Leman, M., & Maes, P. J. (2015). The role of embodiment in the perception of music. *Empirical Musicology Review*, *9*(3–4), 236–246. https://doi.org/10.18061/emr.v913-4.4498

Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends in Neurosciences*, *39*(8), 527–542. https://doi.org/10.1016/j.tins.2016.06.003

Macken, W. J., Taylor, J. C., & Jones, D. M. (2014). Language and short-term memory: The role of perceptual-motor affordance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1257–1270. https://doi.org/10.1037/a0036845

Maes, P.-J., & Leman, M. (2013). The influence of body movements on children's perception of music with an ambiguous expressive character. *PLoS One*, *8*(1), e54682. https://doi.org/10.1371/journal.pone.0054682

Marsh, J. E., Crawford, J. C., Pilgrim, L. K., Sörqvist, P., & Hughes, R. W. (2017). Trouble articulating the right words: Evidence for a response-exclusion account of distraction during semantic fluency. *Scandinavian Journal of Psychology*, *58*(5), 367–372. https://doi.org/10.1111/sjop.12386

Marsh, J. E., Hughes, R. W., & Jones, D. M. (2008). Auditory distraction in semantic memory: A process-based approach. *Journal of Memory and Language*, *58*(3), 682–700. https://doi.org/10.1016/j.jml.2007.05.002

Marsh, J. E., Hughes, R. W., & Jones, D. M. (2009). Interference-by-process, not content, determines semantic auditory distraction. *Cognition*, *110*(1), 23–38. https://doi.org/10.1016/j.cognition.2008.08.003

Marsh, J. E., & Jones, D. M. (2010). Cross modal distraction by background speech: What role for meaning? *Noise & Health*, *12*(49), 210–216. https://doi.org/10.4103/1463-1741.70499

Marsh, J. E., Perham, N., Sörqvist, P., & Jones, D. M. (2014). Boundaries of semantic distraction: Dominance and lexicality act at retrieval. *Memory & Cognition*, *42*(8), 1285–1301. https://doi.org/10.3758/s13421-014-0438-6

Marsh, J. E., Sörqvist, P., Beaman, C. P., & Jones, D. M. (2013). Auditory distraction eliminates retrieval induced forgetting: Implications for the processing of unattended sound. *Experimental Psychology*, *60*(5), 368–375. https://doi.org/10.1027/1618-3169/a000210

Marsh, J. E., Threadgold, E., Barker, M. E., Litchfield, D., Degno, F., & Ball, L. (2021). The susceptibility of compound remote associate problems to disruption by irrelevant sound: A window onto the component processes underpinning creative cognition? *Journal of Cognitive Psychology*, *33*(6–7), 793–822. https://doi.org/10.1080/20445911.2021.1900201

Marsh, J. E., Vachon, F., Sörqvist, P., Marsja, E., Röer, J. P., Richardson, B. H., & Ljungberg, J. K. (2023). Irrelevant changing-state vibrotactile stimuli disrupt verbal serial recall: Implications for theories of interference in short-term memory. *Journal of Cognitive Psychology*. https://doi.org/10.1080/20445911.2023.2198065. ISSN 20-44-5911, E-ISSN 2044-592X.

Marsh, J. E., Yang, J., Qualter, P., Richardson, C., Perham, N., Vachon, F., & Hughes, R. W. (2018). Post-categorical auditory distraction in serial short-term memory: Insights from increased task load and task type. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(6), 882–897. https://doi.org/10.1037/xlm0000492

Martin, R. C., Wogalter, M. S., & Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, *27*(4), 382–398. https://doi.org/10.1016/0749-596X(88)90063-0

McCorkell, N. (2012). Is speech special? The impact of familiarity with irrelevant background sound on working memory. https://e-space.mmu.ac.uk

Meng, Z., Lan, Z., Yan, G., Marsh, J. E., & Liversedge, S. P. (2020). Task demands modulate the effects of speech on text processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(10), 1892–1905. https://doi.org/10.1037/xlm0000861

Miranda, R. A., & Ullman, M. T. (2007). Double dissociation between rules and memory in music: An event-related potential study. *NeuroImage*, *38*(2), 331–345. https://doi.org/10.1016/j.neuroimage.2007.07.034

Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory & Cognition*, *44*(6), 922–936. https://doi.org/10.3758/s13421-016-0608-9

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269. https://doi.org/10.3758/BF03213879

Neath, I. (2000). Modelling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, *7*(3), 403–423. https://doi.org/10.3758/BF03214356

Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 361–394). Lawrence Erlbaum Associates.

Neumann, O. (1996). Theories of attention. In O. Neumann & A. F. Sanders (Eds.), *Handbook of perception and action* (Vol. 3, pp. 389–446). Academic Press.

Nittono, H. (1997). Background instrumental music and serial recall. *Perceptual and Motor Skills*, *84*(3_suppl), 1307–1313. https://doi.org/10.2466/pms.1997.84.3c.1307

Oberauer, K., & Lange, E. B. (2008). Interference in working memory: Distinguishing similarity-based confusion, feature overwriting, and feature mitigation. *Journal of Memory and Language*, *58*(3), 730–745. https://doi.org/10.1016/j.jml.2007.09.006

Özdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage*, *33*(2), 628–635. https://doi.org/10.1016/j.neuroimage.2006.07.013

Parmentier, F. B. R., Pacheco-Unguetti, A. P., & Valero, S. (2018). Food words distract the hungry: Evidence of involuntary semantic processing of task-irrelevant but biologically-relevant unexpected auditory words. *PLoS One*, *13*(1), e0190644. https://doi.org/10.1371/journal.pone.0190644

Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory & Cognition*, *20*(3), 314–320. https://doi.org/10.3758/BF03199668

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691. https://doi.org/10.1038/nn1083

Peretz, I., Gosselin, S., Belin, P., Zatorre, R. J., Plailly, J., & Tillmann, B. (2009). Music lexical networks: The cortical organisation of music recognition. *Annals of the New York Academy of Sciences*, *1169*(1), 256–265. https://doi.org/10.1111/j.1749-6632.2009.04557.x

Peretz, I., Kolinsk, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, *117*(6), 1283–1301. https://doi.org/10.1093/brain/117.6.1283

Peretz, I., & Zatorre, R. J. (2005). Brain organisation for music processing. *Annual Review of Psychology*, *56*(1), 89–114. https://doi.org/10.1146/annurev.psych.56.091103.070225

Perham, N., & Currie, H. (2014). Does listening to preferred music improve Reading comprehension performance? *Applied Cognitive Psychology*, *28*(2), 279–284. https://doi.org/10.1002/acp.2994

Perham, N., Hodgetts, H., & Banbury, S. (2013). Mental arithmetic and non-speech office noise: An exploration of interference-by-content. *Noise and Health*, *15*(62), 73–78. https://doi.org/10.4103/1463-1741.107160

Perham, N., & Sykora, M. (2012). Disliked music can be better for performance than liked music. *Applied Cognitive Psychology*, *26*(4), 550–554. https://doi.org/10.1002/acp.2826

Perham, N., & Vizard, J. (2011). Can preference for background music mediate the irrelevant sound effect? *Applied Cognitive Psychology*, *25*(4), 625–631. https://doi.org/10.1002/acp.1731

Perham, N., & Whitney, T. (2012). Liked music increases spatial rotation performance regardless of tempo. *Current Psychology*, *31*(2), 168–181. https://doi.org/10.1007/s12144-012-9141-6

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychonomic Bulletin & Review*, *20*(2), 310–317. https://doi.org/10.3758/s13423-012-0344-5

Pring, L., & Walker, J. (1994). The effects of unvocalized music on short-term memory. *Current Psychology*, *13*(2), 165–171. https://doi.org/10.1007/BF02686799

Reisberg, D., Smith, J. D., Baxter, D., & Sonenshine, M. (1989). "Enacted" auditory images are ambiguous; "pure" auditory images are not. *Quarterly Journal of Experimental Psychology*, *41*(3), 619–641. https://doi.org/10.1080/14640748908402385

Röer, J. P., Bell, R., & Buchner, A. (2013). Self-relevance increases the irrelevant speech effect: Attentional disruption by one's own name. *Journal of Cognitive Psychology*, *25*(8), 925–931. https://doi.org/10.1080/20445911.2013.828063

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behaviour*, *21*(2), 150–164. https://doi.org/10.1016/S0022-5371(82)90521-7

Salamé, P., & Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology, Section*

A, *41*(1), 107–122. https://doi.org/10.1080/14640748908402355

Samson, S., & Zatorre, R. J. (1991). Recognition memory for text and melody of songs after unilateral temporal lobe lesion: Evidence for dual-encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 793–804. https://doi.org/10.1037/0278-7393.17.4.793

Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science*, *14*(6), 317–320. https://doi.org/10.1111/j.0963-7214.2005.00389.x

Schendel, Z. A., & Palmer, C. (2007). Suppression effects on musical and verbal memory. *Memory & Cognition*, *35*(4), 640–650. https://doi.org/10.3758/BF03193302

Schlittmeier, S. J., Hellbrück, J., & Klatte, M. (2008). Does irrelevant music cause an irrelevant sound effect for auditory items? *European Journal of Cognitive Psychology*, *20*(2), 252–271. https://doi.org/10.1080/09541440701427838

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésane, C., Anton, J.-L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage*, *51*(5), 450–461. https://doi.org/10.1016/j.neuroimage.2010.02.023

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349. https://doi.org/10.1111/1469-8986.00172x

Schubotz, R. I., Friederici, A. D., & von Cramon, D. Y. (2000). Time perception and motor timing: A common cortical and subcortical basis revealed by fMRI. *NeuroImage*, *11*(1), 1–12. https://doi.org/10.1006/nimg.1999.0514

Schulze, K., Muller, K., & Koelsch, S. (2011). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *European Journal of Neuroscience*, *33*(1), 189–196. https://doi.org/10.111/j.1460-9568.2010.07470.x

Schutz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: Action-induced modulations of perception. *Trends in Cognitive Sciences*, *11*(8), 349–355. https://doi.org/10.1016/j.tics,2007.06.005

Serafine, M. J., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody-text integration in memory for songs. *Journal of Memory and Language*, *25*(2), 123–135. https://doi.org/10.1016/0749-596X(86)90025-2

Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expression of emotion. *Psychological and Cognitive Sciences*, *110*(1), 70–75. https://doi.org/10.1073/pnas.1209023110

Simmons-Stern, N. R., Deason, R. G., Brandler, B. J., Frustcae, B. S., O'Connor, M. K., Ally, B. A., & Budson, A. E. (2012). Music-based memory enhancement in Alzheimer's disease: Promise and limitations. *Neuropsychologia*, *50*(12), 3295–3303. https://doi.org/10.1016/j.neuropsychologia.2012.09.019

Smith, E. E., & Jonides, J. (1998). Neuroimaging analysis of human working memory. *Biological Sciences*, *95*(20), 12061–12068. https://doi.org/10.1073/pnas.95.20.12061

Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, *33*(11), 1433–1454. https://doi.org/10.1016/0028-3932(95)00074-D

Sörqvist, P. (2010). High working memory capacity attenuates the deviation effect but not the changing-state effect: Further support for the duplex-mechanism account of auditory distraction. *Memory & Cognition*, *38*(5), 651–658. https://doi.org/10.3758/MC.38.5.651

Sörqvist, P., Halin, N., & Hygge, S. (2010). Individual differences in susceptibility to the effects of speech on reading comprehension. *Applied Cognitive Psychology*, *24*(1), 67–76. https://doi.org/10.1002/acp.1543

Sörqvist, P., Nöstl, A., & Halin, N. (2012). Disruption of writing processes by the semanticity of background speech. *Scandinavian Journal of Psychology*, *53*(2), 97–102. https://doi.org/10.1111/j.1467-9450.2011.00936.x

Taylor, P. J., & Marsh, J. E. (2017). E-prime (software). In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The international encyclopedia of communication research methods* (pp. 1–3). Wiley.

Thaut, M. H. (2009). The musical brain. An artful biological necessity. *Karger Gazette: Music and Medicine*, *70*, 2–4.

Thompson, L. M., & Yankeelov, M. J. (2012). Music and the phonological loop. Paper presented at the European Society for the Cognitive Science of Music, July 23-28, 2012, Thessaloniki, Greece.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychological Science*, *12*(3), 248–251. https://doi.org/10.1111/1467-9280.00345

Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2011). Fast and loud background music disrupts reading comprehension. *Psychology of Music*, *40*(6), 700–708. https://doi.org/10.1177/0305735611400173

Toth, J. P., & Hunt, R. R. (1999). Not one versus many, but zero versus any: Structure and function in the context of the multiple memory systems debate. In J. K. Foster & M. Jelicic (Eds.), *Memory: Systems, process, or function?* (pp. 232–272). Oxford University Press.

Tremblay, S., & Jones, D. M. (1998). Role of habituation in the irrelevant sound effect: Evidence from the effects of token set size. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 659. https://doi.org/10.1037/0278-7393.24.3.659

Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1750–1754. https://doi.org/10.1037/0278-7393.26.6.1750

Tsang, C. D., Friendly, R. H., & Trainor, L. J. (2011). Singing development as a sensorimotor interaction problem. *Psychomusicology: Music, Mind, and Brain*, *21*(1–2), 31–44. https://doi.org/10.1037/h0094002

Vachon, F., Labonté, K., & Marsh, J. E. (2017). Attentional capture by deviant sounds: A non-contingent form of auditory distraction? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 622–634. https://doi.org/10.1037/xlm0000330

Vasilev, M. R., Kirkby, J. A., & Angele, B. (2020). Auditory distraction during reading: A Baysian meta-analysis of a continuing controversy. *Perspectives on Psychological Science*, *13*(5). https://doi.org/10.1177/1745691617747398

Watkins, M. J., & Allender, L. E. (1987). Inhibiting word generation with word presentations. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, *13*(4), 564–568. https://doi.org/10.1037/0278-7393.13.4.564

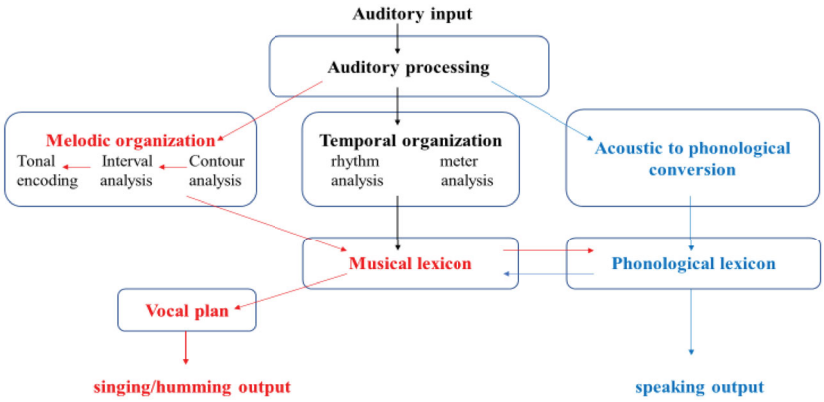
Williamson, V. J., Mitchell, T., Hitch, G., & Baddeley, A. D. (2010). Musicians' memory for verbal and tonal material under conditions of irrelevant sound. *Psychology of Music*, *38*(3), 331–350. https://doi.org/10.1177/0305735609351918

Witt, J. K. (2011). Action's effect on perception. *Current Directions in Psychological Science*, *20*(3), 201–206. https://doi.org/10.1177/0963721411408770

Woodward, A. J., Macken, W. J., & Jones, D. M. (2008). Linguistic familiarity in short-term memory: A role for (co-)articulatory fluency? *Journal of Memory and Language*, *58*(1), 48–65. https://doi.org/10.1016/j.jml.2007.07.002

Appendices

Appendix 1



Cognitive neuropsychological model of processing pathway based on Peretz and Coltheart (2003). Note. melody route spoken-lyrics route.

Appendix 2. Assessment criteria for the measurement of humming retrieval accuracy per bar

Score	0	1	2	3	4
Criteria	inaccurate, no rewardable content	limited accuracy, repeated errors	broadly accurate, some errors	mainly accurate	accurate + fluent
total bars	bars correct	bars correct	bars correct	bars correct	bars correct
8	0	1–2	3–4	5–6	7–8
12	0	1–3	4–6	7–9	10–12
14	0	1–3.5	3.5–7	7–10.5	10.5–14
16–15	0	1–4	5–8	9–12	13–16
20	0	1–5	6–10	11–15	16–20

Assessment criteria for the measurement of spoken lyrics retrieval accuracy per bar

Score	0	1	2	3	4
Criteria	inaccurate, no rewardable content	limited accuracy, repeated errors	broadly accurate, some errors	mainly accurate	accurate + fluent
total bars	bars correct	bars correct	bars correct	bars correct	bars correct
8	0	1–2	3–4	5–6	7–8
12	0	1–3	4–6	7–9	10–12
14	0	1–3.5	3.5–7	7–10.5	10.5–14
16–15	0	1–4	5–8	9–12	13–16
20	0	1–5	6–10	11–15	16–20a