

THE ROLES OF ABSOLUTE PITCH AND TIMBRE IN PLINK PERCEPTION

REBECCA N. FAUBION-TREJO
Virginia Polytechnic Institute and State University

JAMES T. MANTELL
St. Mary's College of Maryland

LISTENERS CAN RECOGNIZE MUSICAL EXCERPTS LESS than one second in duration (plinks). We investigated the roles of timbre and implicit absolute pitch for plink identification, and the time course associated with processing these cues, by measuring listeners' recognition, response time, and recall of original, mistuned, reversed, and temporally shuffled plinks that were extracted from popular song recordings. We hypothesized that performance would be best for the original plinks because their acoustic contents were encoded in long-term memory, but that listeners would also be able to identify the manipulated plinks by extracting dynamic and average spectral content. In accordance with our hypotheses, participants responded most rapidly and accurately for the original plinks, although notably, were capable of recognition and recall across all conditions. Our observation of plink recall in the shuffled condition suggests that temporal orderliness is not necessary for plink perception and instead provides evidence for the role of average spectral content. We interpret our results to suggest that listeners process acoustic absolute pitch and timbre information to identify plinks and we explore the implications for local and global acoustic feature processing.

Received: November 16, 2020, accepted September 10, 2021.

Key words: music memory, plink, absolute pitch, timbre, popular music

AUDITORY SCIENTISTS HAVE UTILIZED BRIEF sound stimuli to investigate the psychological processes that enable humans to rapidly extract acoustic information and perform tasks such as sound identification, speech discrimination, and music perception. Listeners can distinguish vocal and instrumental sounds with signal durations of approximately 10–50 ms (Bigand et al., 2011; Suied et al., 2014) and

identification of such sounds can occur in as little as 250 ms (Agus et al., 2012), even for stimuli with substantial reductions in signal clarity (Isnard et al., 2016). Similarly, Schweinberger and colleagues (1997) demonstrated that participants can distinguish familiar and unfamiliar voices in as little as 250 ms, and Layman and Dowling (2018) showed that listeners could determine whether music contained vocal content with stimuli durations as short as 100 ms.

Plink, an onomatopoeia established by Krumhansl (2010), refers to a brief clip of music, typically with a duration less than one second. Exposure to plinks occurs in everyday life when adjusting the radio dial, or perhaps more commonly while playing any of various “guess the song” games on a smart phone or speaker. Plink research originated as researchers investigated how much information listeners need to classify music by genre (Gjerdingen & Perrott, 1999, as cited in Gjerdingen & Perrott, 2008) and to determine whether listeners could identify songs based on signals that contained absolute (e.g., pitch, timbre) but not relational (e.g., interval) information (Schellenberg et al., 1999). Following Gjerdingen and Perrott, who showed that listeners can accurately identify the genres of plinks with 250 ms duration, Mace et al. (2011) found that listeners can perform such a task with 125 ms plinks. Furthermore, listeners can perceive emotional content following exposure to plinks with durations of 250 ms or less (Bigand et al., 2005; Filipic et al., 2010; Nordström & Laukka, 2019) and their aesthetic ratings of plinks positively correlate with their liking for much longer excerpts (Belfi et al., 2018). Plazak and Huron (2011) investigated the time course of cognitive processes that are performed when a listener hears a musical excerpt and concluded that instrumentation is determined within about 100 ms, genre identification and the presence of voice content are determined within about 400 ms, and other musical features are decoded at longer time scales.

The fact that listeners can so rapidly process general musical and emotional information from plinks is impressive, but it is perhaps not as surprising as listeners' abilities to recall a plink by name. Schellenberg and colleagues (1999) presented 100 and 200 ms popular (Pop) music plinks to listeners, following a pretest familiarization with 20 s excerpts from each song.

Participants were tasked with matching each plink with the song and artist name from a list. Schellenberg et al. manipulated the timbre of the plinks to determine which acoustic properties are relevant for song recognition. Participants were able to match original and high-pass plinks to their song and artist in as little as 100 ms, but matching performance suffered in the reversed and low-pass filtered conditions. Krumhansl (2010) expanded on Schellenberg et al.'s work with a new plink paradigm. Rather than matching plink names to a list, Krumhansl presented 300 and 400 ms Pop song plinks and asked listeners to report the artist, song title, and their confidence, among other attributes including musical style, emotion, and decade of release. Remarkably, participants recognized more than 25% of the 400 ms plinks, though performance was poorer for 300 ms plinks, and artist recall was better than song recall for both duration conditions. Taken together, Schellenberg et al. showed that listeners can match very brief plinks to their titles while Krumhansl demonstrated that listeners can reliably recall song and artist names for plinks of less than one half second in duration.

The fact that listeners can recognize and recall detailed song identity information from limited and modified acoustic stimuli suggests that typical listeners maintain a vast store of song information in long-term memory (Krumhansl, 2010; Schellenberg et al., 1999). Individuals' song memories contain detailed representations of melodic and rhythmic characteristics, and researchers have shown that normal people maintain implicit memories for absolute tempo (Levitin & Cook, 1996) and pitch (Levitin, 1994; Schellenberg & Trehub, 2003). Implicit memories differ from explicit memories in that, although musically untrained individuals might not be able to identify a given key or tempo, they are able to identify canonical pitches (Ben-Haim et al., 2014; Levitin & Rogers, 2005). Thus, it seems reasonable to hypothesize that implicit knowledge of absolute acoustic information may contribute to plink memory. However, Schellenberg et al. (1999) argued, based on the facts that their plinks were too brief to present tempo or relative pitch information and that listeners recognized reversed plinks less accurately than regular plinks (despite them sharing acoustic features related to absolute pitch), that listeners must have utilized spectral timbral cues to identify the plinks. Siedenbueg and Müllensiefen (2017) developed statistical models to determine which acoustic factors enable listeners to categorize unfamiliar plinks. The authors proposed that multiple timbre-related acoustic cues contribute to listeners' abilities to judge plink similarity. Most recently, Thiesen et al. (2020) created a plink identification

prediction model and found that plink duration, song choice (e.g., some songs are more recognizable than others), and song section were the most important variables related to song recall. With respect to acoustic factors, their model revealed that spectral entropy was the key predictor, but that voice content was also related to song recognition. The authors concluded that plink exposure leads to several simultaneous, independent perceptual-cognitive processes that are ultimately integrated to permit song recall.

In this experiment, we explored the psychological processes that listeners use to decode plinks. The purpose of our work was twofold. First, we set out to replicate the finding that listeners can recognize and recall song and artist names from brief musical excerpts (Krumhansl, 2010; McKellar & Cohen, 2015; Schellenberg et al., 1999). Second, we sought to expand the literature about how absolute pitch and timbre cues contribute to listeners' abilities to recognize plinks. To achieve these goals, we compared listeners' recognition and recall performance for plinks in original, mistuned, and timbre-manipulated conditions to determine the relative weight of importance of each of these acoustic cues. We also measured participants' recognition response times because we hoped to learn more about the time course (Plazak & Huron, 2011) by which listeners access long-term memories including absolute pitch (Levitin, 1994; Schellenberg & Trehub, 2003) and timbre (Schellenberg et al., 1999).

We combined Schellenberg et al.'s (1999) stimulus manipulation approach with Krumhansl's (2010) plink perception procedure by presenting constant duration plinks that varied in their absolute pitch and timbre content. Instead of asking our participants to match plinks to song and artist names from a list (which limits our memory measure to recognition), we measured both recognition and recall for each plink. We explicitly selected highly familiar music for our plink stimuli, and we created plinks from highly recognizable sections of each song, so that we could maximize recognition and recall performance for our stimulus set. This was appropriate for our work because our goal was not to demonstrate the effects of plink duration but instead to show how perception is influenced by the acoustic information presented within each plink. We utilized a between-subjects approach to compare performance across multiple acoustic conditions that were based on the same set of songs. It was important for us to reduce the possibility of floor effects to ensure that any differences in performance across conditions were due to the relevant acoustic manipulations instead of the overall familiarity of the songs that we used.

In addition to presenting plinks in their original form (i.e., 400 ms excerpts from canonical studio recordings), we also presented *reversed* plinks, following Schellenberg et al.'s (1999) logic that reversed stimuli present the same static, absolute pitch and spectral information as the original plinks but different dynamic spectral information. In other words, the sequence of acoustic events present in the original plinks matches the sequence of events stored in long-term memory, whereas the acoustic information in the reversed plinks unfolds over time in a way that does not match memory. However, it seems possible that listeners could match adjacent acoustic events, regardless of their order, to acoustic sequences stored in long-term memory. For example, assuming that a listener can extract acoustic events from either the original (events A B C D E) or reversed (events E D C B A) plinks, then both of these stimuli could be matched to a song stored in long-term memory. Although it is more difficult to recall a backwards melody, it is not impossible (Schulkind, 2004; Schulze et al., 2011; White, 1960). Perhaps a similar recognition mechanism could be leveraged at much shorter time scales in which the unit of analysis is not a note but a much briefer acoustic segment such as an event transition.

Although Schellenberg et al. (1999) demonstrated the importance of dynamic spectral information for plink perception, it may be possible that a listener could match a plink's average spectrum to the spectral information of an artist or song that has been encoded in memory. This strategy could be the basis for listeners' abilities to reliably identify plink genres (Gjerdingen & Perrott, 2008; Krumhansl, 2010; Mace et al., 2011). Alternatively, it may be possible for a listener to match a plink's fine-grained, transient spectral events to a record of those events stored in memory. Following Biederman's (1987) recognition by components model of visual perception, Layman and Dowling (2018) hypothesized that low level spectral features associated with voices and instrumentation could serve as the building blocks that ultimately enable a listener to build up perception of an artist or song.

The distinction between the spectral average and spectral event plink identification strategies is analogous to the debate over the role of local and global features in perception (e.g., Bigand et al., 2009). To investigate the degree to which listeners utilize local and global acoustic features in plink perception, we devised a second timbre manipulation that maintains the average (global) spectral properties of the original and reversed plinks but that eliminates the local (dynamic) temporal cues such as event transitions. Our goal was to create plinks that

could be identified by their average (global) spectral content only. In other words, we intended to smear each plink's spectral content across its duration. Just as researchers have utilized timbre smearing approaches to investigate intelligibility of degraded speech (e.g., Alexander et al., 2011; Boothroyd et al., 1996; Hou & Pavolik, 1994), we sought to investigate the intelligibility of degraded plinks. To this end, we created *shuffled* plinks that randomized the order of acoustic events within each plink.

We also created *mistuned* plinks to determine whether listeners utilize absolute pitch information during plink perception. Schellenberg and Trehub (2003) demonstrated that typical listeners can distinguish between 5 s excerpts in their original and mistuned versions, even for one semitone key changes. Critically, however, listeners could only perform this task for music they were already familiar with, indicating that listeners maintain implicit absolute pitch memories. We adapted Schellenberg and Trehub's technique to create mistuned plinks. We reasoned that if listeners use absolute pitch information during plink perception, then we would observe a difference in recognition and recall between the original and mistuned versions.

We hypothesized that recognition, recall, and response time (RT) would be best for the original condition, as the acoustic information would most accurately match what is stored in long-term memory. We suspected that performance would be the worst in the timbre manipulation conditions (reversed and shuffled) because these manipulated acoustic signals are not encoded in participants' long-term memories. In the mistuned condition, the dynamic spectral content would match long-term memory although tuning would not. However, just as most listeners can easily identify songs performed in different keys, we expected participants to identify mistuned plinks (this was also predicted by Schellenberg et al., 1999, p. 646). Notably, if listeners attempt to match absolute pitch information from the plinks to their implicit absolute pitch memories, then performance of either recognition, recall, or RT may suffer for mistuned plinks compared to original plinks. Alternatively, if absolute pitch access takes time to build up, and plink perception is driven almost entirely by timbre information, then we should not observe performance differences between the original and mistuned plinks. Further, if there is a difference in the time course of perceptual processing for timbre and absolute pitch content, then this difference may emerge as an RT advantage in either the mistuned, reversed, or shuffled conditions. Faster RT in one condition over another may suggest that the critical

acoustic information has been processed (i.e., compared to long-term memories) more rapidly. Given that our shuffled plinks are more acoustically dissimilar to the original plinks than either the reversed or mistuned plinks, we expected them to produce the poorest performance overall. However, based on the hypothesis that listeners extract each plink's average spectrum, we predicted that participants would recognize shuffled plinks at an above-zero level of performance. Additionally, given that artists tend to present similar acoustic arrangements across their songs (at least within an album; Kim et al., 2006), we expected that recall accuracy would be better for artist responses than song responses.

Method

PARTICIPANTS

This research was approved by the St. Mary's College of Maryland (SMCM) IRB before data collection began. We recruited 146 participants through the Psychology Department's research participation pool but we lost one participant's task data due to a computer crash. Due to a logistical error, we failed to collect questionnaire data from one other participant before their debriefing. Of the 145 participants for which we collected questionnaire data, they included 98 (67.59%) female participants, 46 (31.72%) male participants, and one participant (0.69%) who did not report their sex. Participant ages ranged from 17 to 24 ($M = 18.81$, $SD = 1.16$). The three participants who were under 18 provided parental consent prior to participation. Five participants reported a country of birth other than the United States (Australia, Cameroon, Germany, Greece, and Italy) and one participant did not report their country of birth. One hundred thirty-nine participants (95.86%) reported English as their native language, two participants (1.38%) reported Spanish, one participant (0.69%) reported Greek, one participant (0.69%) reported French, one participant (0.69%) reported Tagalog, and one participant (0.69%) did not identify their native language. All participants reported English to be their most comfortable spoken language at the point of participation, except for one participant who did not answer the question. Of the 144 participants who reported their confidence in having absolute pitch, eighty-four participants (58.33%) reported "definitely not," thirty-one (21.53%) reported the value between "definitely not" and "perhaps a little," twenty-four (16.67%) reported "perhaps a little," four participants (2.78%) reported between "perhaps and little" and "yes, absolutely," and one participant (0.69%) reported "yes,

absolutely." Participants reported listening to, on average, 32.72 ($n = 134$, $SD = 29.92$) hours of music per week (12 participants did not provide a response or provided impossible values). Participants were pseudorandomly assigned to one of four auditory conditions: original ($n = 43$), mistuned ($n = 26$), reversed ($n = 30$), or shuffled ($n = 46$).

MATERIALS AND MEASURES

Plink Stimuli

It was important for us to choose recognizable music so that we could minimize the possibility of floor effects while comparing performance across acoustic conditions. Based on findings that the music individuals are exposed to during adolescence and early adulthood maintains a salient position in memory (e.g., Krumhansl & Zupnick, 2013), we selected highly recognizable Pop songs that charted during the time individuals within our population of interest (undergraduate students) attended middle school, high school, and college. We accordingly set our song release date timeframe from 2011 to 2018, targeting the Pop music we expected to have been playing across mainstream media during these years. First, we identified songs through an objective comparison of song popularity. We created an initial list of songs from multiple lists including "Billboard: Best Catchy Songs," "Billboard: Top Ten Summer Songs of The Year," and Grammy Awards Major Nominee of The Year. Next, directed research students, including author RNF-T, narrowed the initial song list to 50 hit songs (listed in Table 1) that were expected to be highly recognizable to students in the sampling population and we downloaded the songs from Amazon.com as .mp3 files. We converted each song to mono with Praat (Boersma & Weenink, 2018). Next, we extracted plinks from the original songs (example stimuli appear in Figure 1). Thiesen and colleagues (2019) pointed out that the position of the excerpt within the song can have dramatic effects on its ability to activate a listener's memory. For example, Krumhansl (2010) created 300 and 400 ms plinks from chorus and non-chorus sections of her songs and found that participants recalled artist and song names more accurately in the 400 ms chorus condition (Thiesen et al., 2020, confirmed the chorus advantage). We imposed an additional rule to enhance overall plink recall: our plinks included spoken or sung vocal content. Krumhansl found that recall was better for plinks that contained word content from the song title than plinks that did not include such information (although Schellenberg et al., 1999, did not find a matching advantage for plinks with vocal content). Our plinks included vocal content but the vocal content

TABLE 1. Descriptive Statistics by Plink

Song	Artist	Released	Recog	Recognition				RT				Recall			
				Orig	Mist	Rev	Shuf	Orig	Mist	Rev	Shuf	Orig	Mist	Rev	Shuf
Bodak Yellow	Cardi B	2017	.50	.77	.31	.60	.30	1.14	2.35	2.60	1.79	.42	.08	.23	.22
Hello	Adele	2015	.43	.58	.38	.57	.24	1.69	2.35	2.03	3.29	.16	.15	.17	.04
Chandelier	Sia	2014	.41	.51	.15	.60	.33	1.50	2.00	1.19	2.04	.35	.08	.43	.15
Bad Blood	Taylor Swift	2014	.39	.84	.42	.27	.02	0.72	1.64	1.92	2.82	.72	.35	.07	.00
Turn Down For What	DJ Snake / Lil Jon	2013	.37	.79	.58	.13	.02	2.02	1.92	3.17	0.77	.07	.04	.03	.00
One Dance	Drake	2016	.37	.53	.27	.30	.33	1.82	3.91	1.52	1.91	.23	.08	.03	.11
Work	Rihanna	2016	.35	.53	.19	.27	.33	1.33	3.73	1.64	1.76	.30	.04	.00	.20
Can't Hold Us	Macklemore / Ryan Lewis	2012	.34	.67	.12	.33	.15	3.57	3.59	5.44	4.80	.30	.04	.13	.04
Rockstar	Post Malone	2017	.31	.70	.23	.10	.13	1.13	2.07	4.67	2.69	.49	.04	.07	.02
All of Me	John Legend	2013	.29	.44	.35	.30	.11	1.94	2.80	3.69	2.40	.33	.19	.17	.00
Blurred Lines	Robin Thicke / Pharrell Williams	2013	.28	.51	.50	.20	.00	2.02	4.13	3.74		.26	.12	.00	.00
Uptown Funk	Mark Ronson / Bruno Mars	2015	.26	.65	.12	.13	.07	1.30	1.35	4.52	1.59	.19	.04	.00	.00
Cheap Thrills	Sia	2016	.26	.19	.15	.53	.20	2.18	4.14	3.01	3.81	.09	.00	.07	.00
1-800-273-8255	Logic	2017	.25	.56	.35	.10	.00	1.72	1.25	2.44		.37	.19	.00	.00
Cups	Anna Kendrick	2012	.25	.56	.23	.10	.07	2.54	1.87	4.47	5.95	.23	.12	.00	.00
Watch Me	Silentó	2015	.23	.53	.19	.10	.04	1.59	1.55	4.56	1.26	.02	.04	.00	.00
Happy	Pharrell Williams	2013	.22	.51	.15	.20	.00	1.62	1.12	2.22		.33	.08	.10	.00
Dark Horse	Katy Perry	2013	.21	.49	.23	.13	.00	1.28	2.99	3.04		.26	.08	.13	.00
Love Yourself	Justin Bieber	2015	.21	.37	.15	.20	.09	2.70	3.08	2.54	1.98	.16	.04	.03	.02
Locked Out of Heaven	Bruno Mars	2012	.19	.42	.19	.10	.04	1.45	2.32	2.82	1.95	.12	.00	.00	.00
Cruise	Florida Georgia Line	2012	.19	.33	.15	.17	.09	2.53	2.98	2.38	2.61	.12	.04	.07	.00
Roar	Katy Perry	2013	.18	.35	.19	.13	.04	1.77	1.87	5.64	2.85	.09	.00	.03	.00
Get Lucky	Daft Punk	2013	.17	.44	.15	.03	.02	1.47	1.33	7.92	4.97	.16	.04	.00	.00
Call Me Maybe	Carly Rae Jepsen	2012	.17	.44	.12	.10	.00	0.97	1.53	4.67		.26	.08	.00	.00
Royals	Lorde	2013	.15	.16	.15	.20	.11	1.79	1.78	1.70	4.08	.07	.00	.00	.00
All About That Bass	Meghan Trainor	2014	.14	.37	.12	.07	.00	1.19	8.01	4.36		.12	.04	.00	.00
Gangnam Style	PSY	2012	.12	.09	.19	.13	.11	0.81	4.04	3.13	3.65	.09	.00	.00	.04
Take Me to Church	Hozier	2014	.11	.19	.12	.10	.04	1.81	2.34	2.58	4.88	.12	.04	.07	.00
Thrift Shop	Macklemore / Ryan Lewis	2012	.11	.07	.08	.30	.04	5.93	3.42	2.48	3.52	.05	.00	.07	.00
Sorry	Justin Bieber	2015	.10	.21	.15	.00	.04	1.28	1.40		1.51	.07	.00	.00	.00
Can't Feel My Face	The Weeknd	2015	.10	.21	.04	.07	.04	2.30	1.00	3.93	0.50	.09	.00	.00	.02
Shake It Off	Taylor Swift	2014	.10	.19	.04	.03	.09	0.98	2.93	0.32	2.97	.05	.00	.00	.00
Treasure	Bruno Mars	2012	.10	.14	.08	.13	.04	1.20	1.27	2.88	4.88	.05	.00	.00	.00
Somebody That I Used To Know	Gotye	2011	.09	.23	.08	.03	.00	4.32	2.67	2.57		.09	.04	.00	.00
Thinking Out Loud	Ed Sheeran	2014	.09	.19	.04	.03	.07	3.45	7.35	7.82	3.02	.02	.00	.00	.00
Wide Awake	Katy Perry	2012	.09	.14	.15	.10	.00	1.10	1.53	1.57		.00	.00	.00	.00
Despacito	Luis Fonsi	2017	.08	.14	.12	.03	.04	0.85	1.91	9.89	1.83	.00	.00	.00	.00
Look What You Made Me Do	Taylor Swift	2017	.08	.09	.12	.07	.04	5.08	3.25	2.06	1.45	.05	.00	.00	.00
Fancy	Iggy Azalea	2014	.07	.19	.04	.00	.02	4.02	2.77		1.07	.09	.00	.00	.00
Alright	Kendrick Lamar	2015	.06	.07	.08	.03	.04	1.47	3.32	2.30	1.80	.05	.00	.00	.00
Stay With Me	Sam Smith	2014	.05	.12	.00	.07	.00	1.26		9.95		.09	.00	.00	.00

(continued)

TABLE 1. (continued)

Song	Artist	Released	Recog	Recognition				RT				Recall			
				Orig	Mist	Rev	Shuf	Orig	Mist	Rev	Shuf	Orig	Mist	Rev	Shuf
We Are Young	Fun.	2011	.05	.09	.08	.03	.00	1.66	9.69	1.00		.05	.00	.00	.00
Stressed Out	Twenty One Pilots	2015	.05	.07	.00	.10	.02	1.21		3.65	7.60	.07	.00	.03	.00
Can't Stop the Feeling	Justin Timberlake	2016	.04	.09	.08	.00	.00	2.21	2.20			.05	.00	.00	.00
Blank Space	Taylor Swift	2014	.04	.05	.04	.07	.02	0.99	1.80	1.29	1.42	.00	.00	.00	.00
Cheerleader	OMI	2014	.04	.02	.08	.07	.02	1.05	4.23	1.64	3.10	.00	.00	.00	.00
Counting Stars	OneRepublic	2013	.03	.07	.04	.03	.00	6.11	4.05	0.55		.05	.00	.00	.00
Radioactive	Imagine Dragons	2012	.03	.09	.00	.00	.00	2.59				.02	.00	.00	.00
Starships	Nicki Minaj	2012	.01	.05	.00	.00	.00	0.83				.05	.00	.00	.00
We Can't Stop	Miley Cyrus	2013	.01	.00	.00	.03	.00			2.67		.00	.00	.00	.00

Note. Recog = Proportion of listeners across all conditions who recognized each plink; Recognition = Proportion of listeners within each condition who recognized the plink; RT = Mean recognition response time (s) within each condition; Recall = Proportion of listeners within each condition who correctly recalled both the song and artist name. Orig = original condition; Mist = mistuned condition; Rev = reversed condition; Shuf = shuffled condition. Empty cells indicate that no responses satisfied the condition. This table is sorted by most recognized plinks across all conditions.

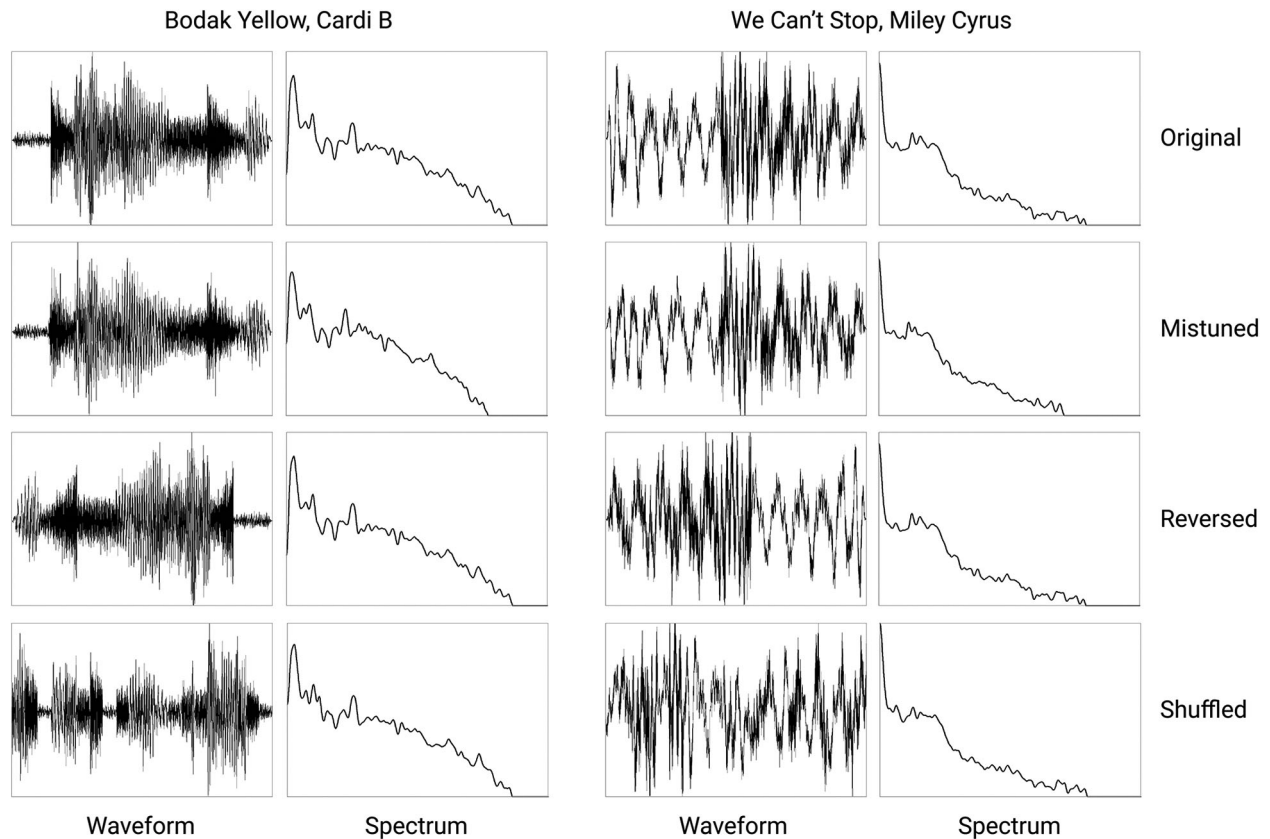


FIGURE 1. Example stimuli. Waveforms and spectra for the most (Bodak Yellow) and least (We Can't Stop) recognized plinks across all conditions. Images drawn in Praat. Bounding boxes around waveforms represent duration (X; 0–400 ms) and amplitude (Y; intensity scaled to 70 dB SPL). Bounding boxes around cepstral smoothed (500 Hz bandwidth) spectra represent frequency (X; 0–20 kHz) and power (Y; 0–40 dB/Hz).

never contained any part of the song title to reduce the possibility that semantic memory could drive memory effects across songs.

We created mistuned copies of each plink in Audacity (Audacity Team, 2018) by applying the “Change Pitch” function (semitones: -2, 2; with “high quality stretching” enabled, which uses a sub-band sinusoidal modeling synthesis algorithm to ensure that signal duration is preserved during pitch shifting). We originally planned to present sharp and flat versions that were mistuned by ± 2 semitones (ST) based on Schellenberg and Trehub’s (2003) finding that individuals can accurately discriminate between original and 2 ST mistuned excerpts of familiar songs. However, to reduce the number of stimulus combinations in our study, we ultimately decided to only use the flat-tuned versions. Henceforth we refer to the -2 ST stimuli as mistuned.

Next, we created our timbre-modified plinks. We created reversed plinks by reversing the original plinks in Praat. The reversed and original plinks contained identical average spectral and pitch information but dissimilar fine-grained spectrotemporal information (Schellenberg et al., 1999). To create the shuffled plinks, we first considered procedures employed by other researchers. Levitin and Menon (2005) used scrambled classical music stimuli in an fMRI investigation of the brain bases of fine-grained, local feature processing in music perception. They scrambled their music stimuli by dividing long, 23 s excerpts into 250–350 ms segments and then randomly concatenated the segments with crossfade. Bigand and colleagues (2009) examined the role of musical feature stability for familiar and unfamiliar music by extracting and concatenating non-adjacent segments into a new signal. In their work, segments varied in duration between 250–850 ms (with fade) for a total signal duration of 11–26 s, depending on the piece. In contrast to these two studies and prior plink research, we presented listeners with shuffled 400 ms plinks. We wrote a Praat script to batch transform the original plinks into temporally shuffled versions according to the following algorithm (see Figure 2). For each original plink, the script divided the excerpt into 20 ms segments (there were 20 segments per 400 ms plink), applied 1 ms fade in/out to each segment (to avoid acoustic artifacts when the segments are combined), concatenated the segments in random order, and saved the shuffled plink. As can be seen in Figure 1, while the original and shuffled plink waveforms were dissimilar, their spectra were very similar (within each song). This indicates that our shuffle manipulation successfully disrupted local, dynamic spectral content while preserving global, average spectral content.

In total we created 50 plinks \times 4 conditions (original, mistuned, reversed, shuffled) = 200 unique plinks. We created two additional stimuli. First, we created a plink from the song “Let It Go” by Idina Menzel to serve as a stimulus for practice trials only. The practice plink matched the characteristics of the plinks in the original condition except that it contained sound from two words of the title (“Let it”) to further enhance recognition. Second, we used Praat to create a 400 ms segment of random Gaussian noise to be used as a control stimulus. The control stimulus served to reduce the possibility of participants guessing the purpose of the study and to enable us to assess participant attention during the experiment (participants should not recognize or recall songs from the control stimulus). By including a stimulus that should not be recognized as a song, we were able to confirm the validity of our recognition measure by ruling out possible misuse of the “yes” response. Finally, we wrote a Praat script to normalize all stimuli (scale RMS intensities to 70 dB SPL; add 10 ms onset and offset ramps to eliminate sound presentation artifacts), and we listened to the stimuli to confirm the sound properties of each plink. Stimuli are available at <https://osf.io/ma589/>.

Plink Perception Measures

We measured plink memory in two ways: recognition and recall. According to the dual-process theory of memory, recognition (identifying familiarity with a stimulus) is a separate process from recall (recollecting the relevant semantic qualities of a stimulus) and they differ in time course, reliability, and neural activity (Hintzman et al., 1998; Mandler, 2008). First, we defined *recognition* as a “Y” keypress after excerpt exposure signaling familiarity to the stimulus. We chose to use a single keypress for recognition so that we could measure participants’ recognition RTs. When we designed the study, we reasoned that a latency measure might be sensitive to the perceptual processes responsible for plink memory across conditions (Hintzman et al., 1998). Thus, for example, we speculated that if one condition presented faster response times than another, it could indicate that the acoustic characteristics of that condition are better able to activate song memory. The disadvantage of the recognition measure is that there is no way to verify response accuracy. For example, a “Y” keypress could indicate legitimate plink recognition, a recognition error (identifying a different song or artist), or a response error (e.g., accidental keypress). We supplemented our recognition measure with another measure of memory. We defined *recall* as recollection of the correct song title and/or artist. The

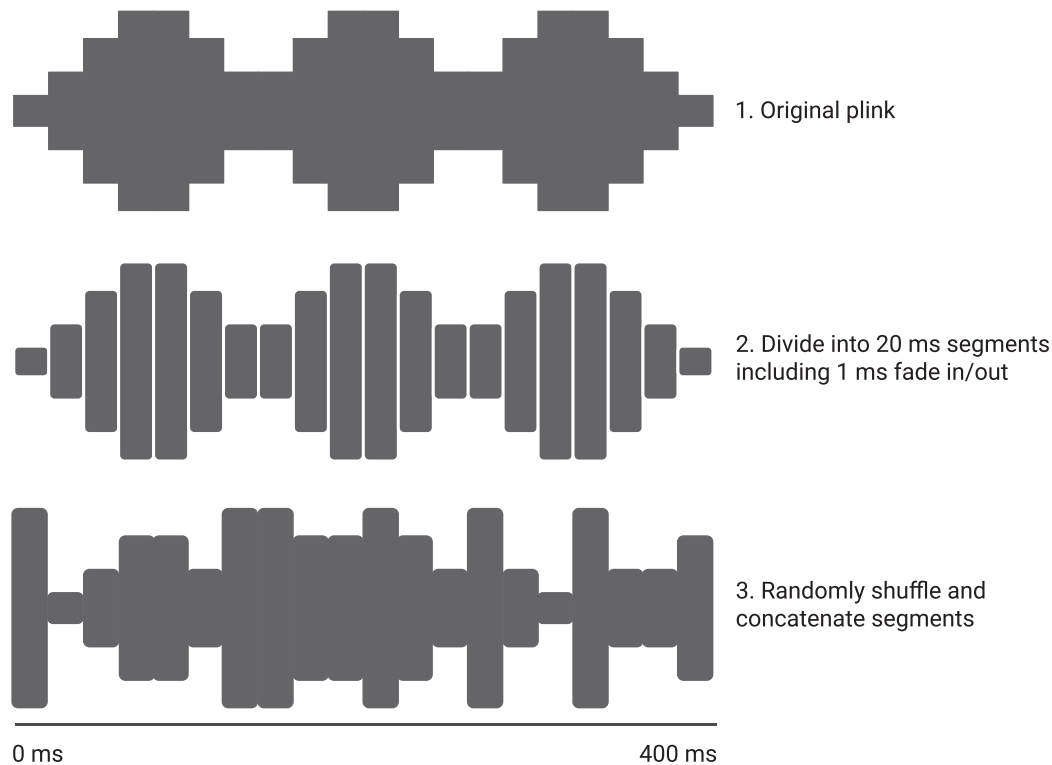


FIGURE 2. Shuffled plink construction. Stylized waveforms represent the construction of a shuffled plink. Rounded segment edges represent fade in/out. Horizontal line represents total duration. The temporal order of the segments in the shuffled plinks did not correlate with recognition or recall performance (see Supplementary Materials accompanying the online version of this paper at mp.ucpress.edu).

major advantage of this measure is that we can verify participants' recall accuracy because we know the song and artist names for the plinks. The disadvantage of the recall measure is that, due to differences in participants' typing speed, it cannot be used to infer anything about the time course of perceptual processing (thus we did not measure recall response time). We separately coded recall accuracy for song, artist, and both because song and artist recall accuracy can vary independently (e.g., if a listener knows the song name but not the artist). Taken together, our measures of recognition, RT, and recall provide a robust assessment of plink memory.

Questionnaires

Participants completed three questionnaires. The general information questionnaire served to gather participants' demographic information including age, gender, handedness, and hearing and spoken language pathology. We also included items to assess participants' subjective comfort during the task and the amount of effort they exerted in completing it. The language background questionnaire assessed participants' place of birth, native language, additional spoken languages, family

language history, and, for nonnative talkers of English, age of English language acquisition and current comfort level in speaking their languages. The music background questionnaire assessed participants' experiences with music including items related to their experiences with formal and informal music education. It asked participants to rate their own singing capabilities and their enjoyment and frequency of singing. The questionnaire also inquired about participants' vocal and instrumental training and performance experiences. Finally, following a printed, traditional definition of absolute pitch, it asked participants to indicate if they have absolute pitch on a five-point scale (1 = *definitely not*; 5 = *definitely yes*).

PROCEDURE

Participants signed up for a 45-minute research session, during which up to five students could participate simultaneously, via an online participation platform. Participants received research participation credits to apply toward a relevant course grade. Data collection occurred in a room containing multiple computer stations that were divided with privacy partitions. Each

station included a Windows PC with widescreen monitor, keyboard, mouse, and a pair of Audio-Technica ATH-M30 circumaural headphones. Following a review of the consent form, the researcher instructed participants to wear their headsets and begin the auditory portion of the experiment (described below). Participants completed the experiment at their own pace. They were asked to silently alert the researcher after they completed the auditory trials so that the researcher could provide a questionnaire packet for silent completion at the workstation. Upon completion, participants were debriefed outside of the experimental room to avoid disturbing other participants.

We programmed stimulus presentation and response collection with PsychoPy (version 1.90.2, Peirce et al., 2019). Throughout the auditory trials, the computer screen displayed a black background and light gray font. The PsychoPy script was loaded at each workstation prior to participants' arrival. The initial screen instructed participants to wait for instructions from the experimenter, who directed participants to press "Enter" after completing informed consent. The next screens reminded participants to wear headphones and displayed instructions including a general description of the task and response options for recognition and recall of song and artist names. After reading the instructions, participants pressed "Enter" to continue to the practice trials (described below).

Each auditory trial utilized the following format. A blank screen displayed for 600 ms followed by on-screen text "Do you know this song?" At 1 s after the start of the trial, the 400 ms plink played three times with 1 s interstimulus intervals. We chose three exposures based on our concern that a single exposure was not sufficient if a participant did not attend to the beginning of the trial. Furthermore, because we were interested in their timed responses, we chose not to allow the participants to control stimulus repetition. At the onset of the third auditory presentation, on-screen text, "press Y (yes) or N (no)," appeared to remind participants about their song recognition response options. The computer was programmed not to accept keypresses until the beginning of the third presentation when the response instructions appeared. At that point, the response timer began so that we could measure recognition response time. If the participant typed "N," the next trial began. If the participant typed "Y," a pop-up window appeared for the participant to type the name of the song and artist and to rate their confidence in their song and/or artist guess by selecting a number between 0 (*no confidence*) and 5 (*maximum confidence*). Participants were not required to enter any responses at

this window and they were informed during the instructions that spelling did not count. When either the "OK" or "Cancel" button was clicked, the popup window disappeared and the next trial began.¹

The practice trials served to acclimatize participants to the brief duration of the plinks and to enable them to practice responding with either a "Y" (yes) or "N" (no) keypress to indicate their recognition of the plink. The practice trials followed a similar format to the typical trials except that additional information appeared to the participant during practice. First, an instruction screen described the structure of the two practice trials. Next, the on-screen text for the first practice trial instructed the participant to respond "N" even if they recognized the song (the computer would not accept any other keypress). Following this "N" keypress, the second practice trial began. This practice trial explicitly displayed the name of the song and artist (Let It Go, Idina Menzel) and informed participants to press "Y" to continue (the computer did not accept any other keypress). Following this "Y" keypress, participants viewed the popup window and practiced entering the artist name, song name, and their confidence. Following the second practice trial, a final instruction screen stated, "The practice trials are over. Now you are ready for the real trials. Do your best to identify the songs. No matter what, please respond as accurately as you can." Next, participants completed 75 randomly ordered trials including 50 plinks and 25 noise control trials. Because we utilized a between-subjects design, each participant heard plinks from just one of the four conditions (original, mistuned, reversed, or shuffled). Upon completion of the main phase of the experiment, a screen instructed participants to tell the experimenter they are finished so that they could begin the questionnaires.

Results

DATA ANALYSES

We chose to utilize nonparametric inferential analyses due to the large, non-normal variation in data across participants and songs (Figure 3; recognition and recall performance is positively skewed across conditions). In other words, some participants recognized many plinks, and some participants did not recognize any. Likewise,

¹ Due to a PsychoPy limitation, the appearance of the popup window caused the main experiment display to be temporarily minimized as the popup window was displayed over the Windows desktop. To eliminate potential distractions, we modified all of the workstations before data collection began to remove program icons and set the color of the Windows desktop background to match the main experiment display.

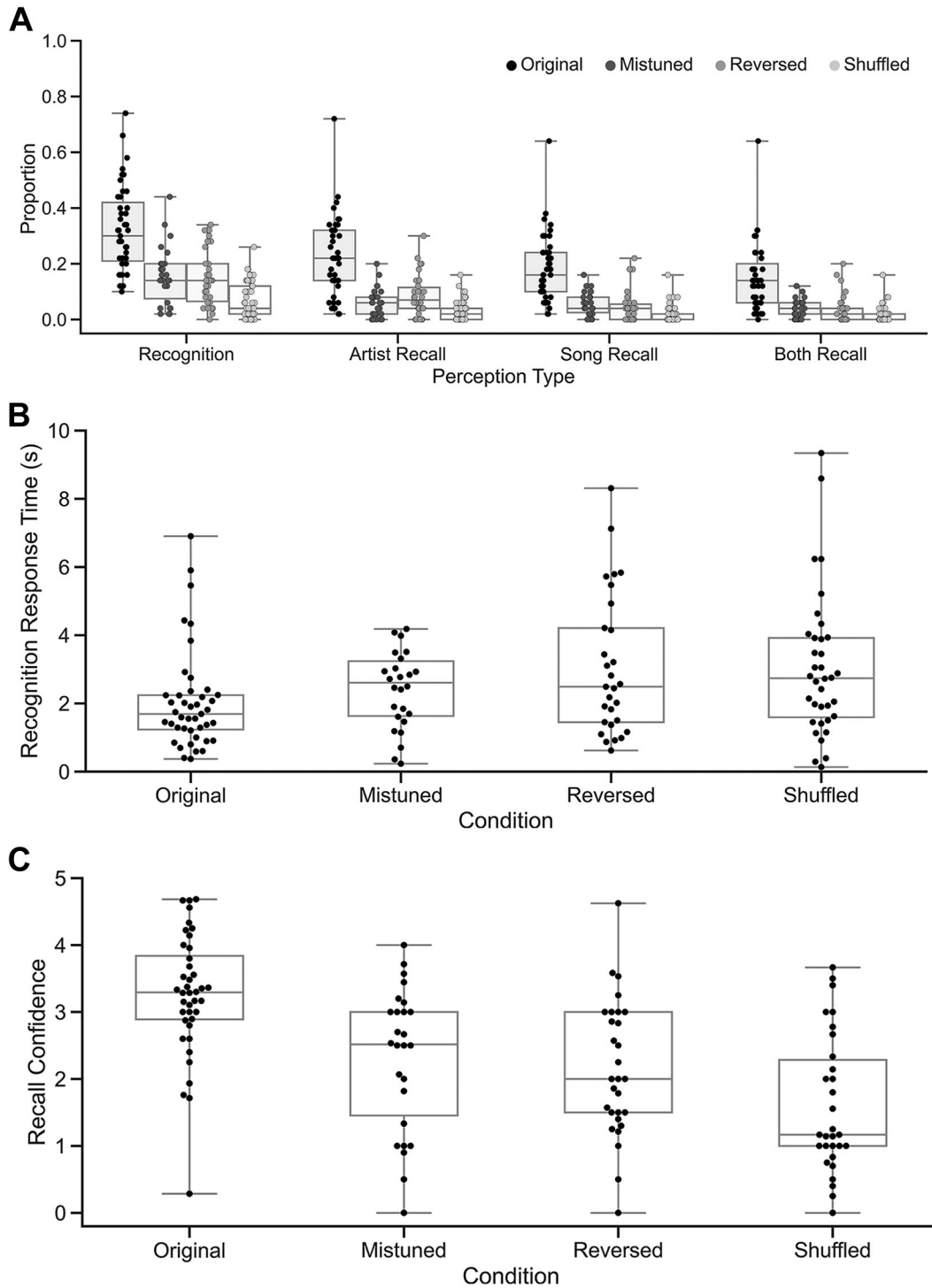


FIGURE 3. Plink recognition, recall, recognition response time, and recall confidence. Data points represent participant means. Box plots depict minimum, maximum, interquartile range, and median. (A) Plink recognition and recall. (B) Recognition response time; a single data point in the mistuned condition (18.15 s) extends beyond the y-axis limits. (C) Recall confidence.

TABLE 2. Frequency of Response Types

	Correct	Exchanged	Partial	SW Error	Incorrect	No Response	Total
Original	390 (461)	16 (9)	12 (5)	59	62 (75)	1,611 (1,600)	2,150
Mistuned	71 (80)	0 (0)	0 (3)	15	46 (82)	1168 (1135)	1300
Reversed	67 (129)	1 (1)	0 (1)	3	57 (62)	1,372 (1,307)	1,500
Shuffled	45 (73)	0 (0)	0 (1)	0	41 (69)	2,214 (2,157)	2,300
Total	573 (743)	17 (10)	12 (10)	77	206 (288)	6,365 (6,199)	7,250

Note. Song (artist) response frequencies for each condition. Partial = Partially Correct. SW Error = Song Word Error (SW Errors only apply to song responses). Condition (row) totals are identical for both song and artist responses. Noise (control) trials are not included. Exchanged responses, but not partially correct responses, were coded as Correct responses in the inferential data analysis.

some plinks were highly recognizable whereas others were never recognized. We report appropriate measures of central tendency (median) and variation (interquartile range; *IQR*) with each of the nonparametric inferential statistics and we report means and standard deviations in the tables. We conducted the inferential analyses in two ways. The *listener analyses* treated participant as a random factor to determine how condition affects plink memory across listeners. We conducted listener analyses with between-subjects Kruskal-Wallis tests and eta squared effect sizes (Cohen, 2013) followed by Dunn-Bonferroni post hoc pairwise comparisons (these include adjusted *p* values to reduce the possibility of false positives). The *song analyses* treated song as a random factor to determine how condition affects plink memory across songs. We conducted song analyses with repeated measures Friedman tests and Kendall's coefficient of concordance (*W*; Cohen, 2013) as effect size, followed by Dunn-Bonferroni post hoc pairwise comparisons. The inferential results are occasionally presented with various sample size values due to missing data (e.g., some participants did not report recall confidence and those data are omitted from their respective analysis). For both the song and listener analyses of recall data, the input data comprises responses we coded as correct and exchanged. We omitted incorrect, song word error, and partially correct responses (see Table 2; for more detailed information on song response coding, see the Supplementary Materials accompanying the online version of this paper at mp.ucpress.edu). For this reason, our results are more conservative than they would have been had we included the partially correct responses. We reasoned that our predictions would be partially supported if they were upheld in either analysis and strongly supported if they were satisfied in both analyses. Due to the

similarity between listener and song results, only listener results have been included in the main text. Please see the Supplementary Materials for song analysis results. Data files are available at <https://osf.io/ma589/>.

RECOGNITION

The Kruskal-Wallis test indicated a statistically significant difference in recognition between the four conditions, $\chi^2(3, n = 145) = 67.75, p < .001, \eta^2 = .46$. Dunn-Bonferroni post hoc tests revealed that the proportion of recognized plinks was higher in the original condition ($Md = .32; IQR = .24$) than the mistuned ($Md = .14, IQR = .15; p = .001$), reversed ($Md = .14, IQR = .16; p < .001$), and shuffled ($Md = .10, IQR = .10; p < .001$) conditions. Recognition was lower for shuffled plinks than mistuned ($p = .006$) and reversed ($p = .010$) plinks. There was no significant difference in recognition between the reversed and mistuned conditions ($p = 1$) (see Figure 3A).

RECOGNITION RT

The Kruskal-Wallis test revealed a statistically significant difference in recognition RT between the four conditions, $\chi^2(3, n = 134) = 9.12, p = .028, \eta^2 = .05$. The post hoc comparison indicated that response time was faster in the original condition ($Md = 1.64$ s, $IQR = 1.11$) than the shuffled condition ($Md = 2.53, IQR = 2.10; p = .050$). There were no other differences in RTs between the original, mistuned ($Md = 2.61, IQR = 1.78$), reversed ($Md = 2.49, IQR = 3.16$), and shuffled condition: original-mistuned ($p = .357$); original-reversed ($p = .107$); mistuned-reversed, mistuned-shuffled, and reversed-shuffled ($p = 1$) (see Figure 3B).

ARTIST RECALL

The Kruskal-Wallis test revealed a statistically significant difference in artist recall between the four

conditions, $\chi^2(3, n = 145) = 63.17, p < .001, \eta^2 = .43$. The proportion of artists recalled was higher in the original condition ($Md = .22, IQR = .20$) than the mistuned ($Md = .06, IQR = .07; p < .001$), reversed ($Md = .08, IQR = .09; p = .001$), and shuffled ($Md = .04, IQR = .05; p < .001$) conditions. Artist recall was higher for reversed than shuffled plinks ($p = .009$), but the difference between mistuned and shuffled conditions was not statistically significant ($p = .256$). The mistuned and reversed conditions did not differ ($p = 1$) (see Figure 3A for song, artist, and both recall across the four conditions).

SONG RECALL

The Kruskal-Wallis test indicated a statistically significant difference in song recall between the four conditions, $\chi^2(3, n = 145) = 75.25, p < .001, \eta^2 = .51$. Post hoc comparisons revealed that the proportion of songs recalled was higher in the original ($Md = .17, IQR = .16$) condition than the mistuned ($Md = .04, IQR = .06; p < .001$), reversed ($Md = .04, IQR = .06; p < .001$), and shuffled ($Md = .02, IQR = .04; p < .001$) conditions. Mistuned song recall was higher than shuffled ($p = .010$) but not reversed ($p = 1$) conditions. Song recall between reversed and shuffled conditions did not statistically differ ($p = .302$).

BOTH RECALL

The Kruskal-Wallis test indicated a statistically significant difference in recall of both song and artist between the four conditions, $\chi^2(3, n = 145) = 59.03, p < .001, \eta^2 = .40$. Recall of both song and artist was higher in the original condition ($Md = .14, IQR = .16$) than in the mistuned ($Md = .04, IQR = .05; p = .001$), reversed ($Md = .02, IQR = .05; p < .001$), and shuffled ($Md = .02, IQR = .04; p < .001$) conditions. Recall was higher for mistuned than shuffled plinks ($p = .045$) whereas the difference between reversed and shuffled was not statistically significant ($p = .450$). Recall did not differ between mistuned and reversed conditions ($p = 1$).

RECALL CONFIDENCE

The Kruskal-Wallis test indicated a statistically significant difference in recall confidence between the four conditions, $\chi^2(3, n = 125) = 37.31, p < .001, \eta^2 = .28$. Confidence was statistically significantly higher in the original condition ($Md = 3.29$, as measured on a scale between 0 and 5, $IQR = 1.04$) than the mistuned ($Md = 2.52, IQR = 1.79; p = .006$), reversed ($Md = 2.00, IQR = 1.55; p < .001$), and shuffled ($Md = 1.17, IQR = 1.46; p < .001$) conditions. The difference between mistuned and reversed conditions did not

differ ($p = 1$), nor did mistuned-shuffled ($p = .171$) or reversed-shuffled ($p = .569$) conditions (see Figure 3C).

ADDITIONAL ANALYSES

Our sample was able to recognize and recall shuffled plinks at an above-zero level of performance, but performance in that condition was lowest overall. Based on an anonymous reviewer's suggestion, we performed a series of one-sample Wilcoxon signed rank tests to test the null hypothesis that the population medians of recognition and recall are zero for shuffled plinks. The results revealed above-zero levels of recognition, $z = 5.25, p < .001, r = .77$, and recall of artist, $z = 4.66, p < .001, r = .69$, recall of song, $z = 4.19, p < .001, r = .62$, and recall of both artist and song, $z = 4.02, p < .001, r = .59$.

As we noted in the Method, one disadvantage for recognition measures of memory is that accuracy of responses cannot be determined because we do not know participants' prior exposure to the songs in the stimulus set. But, if participants are truthfully reporting their recognition of each plink, then recognition should positively correlate with recall. Thus, we investigated the validity of our song recognition measure by calculating Spearman's rank-order correlation between participants' mean proportion of plinks recognized and mean proportion of plinks with both song and artist correctly recalled, $r_s(143) = .79, p < .001$. Though we report the overall correlation here, the correlations between recognition and recall were positive and statistically significant in all four conditions.

Given that we selected popular songs for our stimulus set, we wondered whether our participants' listening habits were related to their performance. We correlated plink recall with participants' self-reported weekly music listening (hours). There was a significant positive correlation between music listening and both artist recall, $r_s(121) = .21, p = .021$, and recognition $r_s(119) = .23, p = .013$. The correlations between weekly listening and song recall, $r_s(119) = .15, p = .090$, or both recall, $r_s(119) = .16, p = .070$, were not significant. We also correlated participants' self-reported absolute pitch ratings and their recall performance, but the correlations were not statistically significant (all p values $> .382$).

Krumhansl (2010) asked her participants to rate their recognition of each plink on a Likert scale and the recognition ratings provided in her study serve a similar function as the confidence responses that we collected. If our confidence ratings correlate with recall ratings, then that would validate the use of recognition rating scales as a singular measure of plink memory. This

TABLE 3. Descriptive Statistics by Condition

Condition	<i>n</i>	Recognition	RT	Song	Artist	Both	Confidence
Original	43	.33 (.16)	2.08 (1.49)	.19 (.12)	.23 (.14)	.15 (.12)	3.26 (0.90)
Mistuned	26	.16 (.10)	2.98 (3.29)	.06 (.04)	.06 (.05)	.04 (.03)	2.31 (1.07)
Reversed	30	.15 (.10)	3.09 (2.06)	.05 (.06)	.09 (.07)	.04 (.05)	2.15 (1.02)
Shuffled	46	.10 (.06)	2.94 (2.13)	.03 (.04)	.05 (.04)	.03 (.04)	1.61 (1.03)

Note. *n* = condition sample size; Recognition = *M* (*SD*) proportion of plinks that were recognized; RT = *M* (*SD*) recognition response time (s); Song/Artist/Both = *M* (*SD*) proportion of correctly recalled songs/artists/both; Confidence = *M* (*SD*) recall confidence between 0 (lowest) and 5 (highest).

could be a tempting strategy because it is much easier to collect plink recognition ratings than collecting and coding recall responses. The Spearman rank-order correlation between participants' mean confidence ratings and mean proportion of plinks for which both song and artist were correctly recalled was positive and significant, $r_s(123) = .50, p < .001$. However, when we examined this correlation across all four conditions, it was only significant within the original condition, $r_s(38) = .34, p = .030$. This result suggests that recognition rating scales could serve as a valid predictor of recall, but only for unaltered plinks.

Does response time predict response accuracy? In other words, is there a relationship between response time and recognition or recall? We examined this question by first conducting a Spearman correlation between participants' mean response time and mean proportion of plinks recognized. The result was negative and significant, $r_s(132) = -.29, p = .001$, indicating that lower overall response times were associated with higher recognition. However, when we examined this correlation within each condition, the only significant association we found was in the original condition, $r_s(41) = -.38, p = .01$ (all other *p* values $> .10$). Next, we examined the Spearman correlation between mean response time and mean proportion of plinks where both song and artist were correctly recalled. The result was $r_s(132) = -.28, p = .001$, indicating that the participants who entered their recognition responses more rapidly also recalled the plinks more accurately.

To test our prediction that artist recall would be better than song recall, we conducted a series of Wilcoxon signed rank tests on the listener data to compare artist and song recall within each condition. The results revealed that artist recall was statistically significantly better than song recall in the original ($z = 3.27, n = 43, p = .001, r = .50$), reversed ($z = 4.10, n = 30, p < .001, r = .75$), and shuffled ($z = 3.57, n = 46, p < .001, r = .53$) conditions, but not in the mistuned condition ($z = .90, n = 26, p = .37, r = .18$). We also examined the relationship between artist recall and song recall. We expected them to be positively correlated, and indeed, the Spearman correlation between participants' mean

proportions of artists and songs recalled was positive and statistically significant, $r_s(143) = .87, p < .001$ (this relationship held true within all four conditions; $p_s < .001$). Based on condition means (see Table 3), we expected that artist recall would be higher than song recall overall, but we wondered whether the relative proportions of recalled artists and songs differed across conditions. To examine this possibility, we calculated the *artist recall ratio* for each participant as the ratio of recalled artists to recalled songs. If a participant recalled the same number of artists and songs, the artist recall ratio equals 1 (whereas larger values represent more artists than songs recalled and smaller values represent fewer artists than songs recalled). The artist recall ratio cannot be computed for participants who recalled zero artists and/or zero songs. Thus, 38/145 participants were removed from the overall analysis (original $n = 0/43$; mistuned $n = 4/26$; reversed $n = 10/30$; shuffled $n = 24/46$). Next, we submitted participants' artist recall ratios to a between-subjects Kruskal-Wallis test that revealed a statistically significant difference in artist recall ratio between the four conditions, $\chi^2(3, n = 107) = 12.72, p = .005, \eta^2 = .09$ (original $Md = 1.13, IQR = .73$; mistuned $Md = 1.00, IQR = .83$; reversed $Md = 2.00, IQR = 1.80$; shuffled $Md = 1.13, IQR = 1.00$). Dunn-Bonferroni post hoc tests revealed that the artist recall ratio was significantly higher, thus indicating relatively more artists than songs recalled, in the reversed condition than in the original condition ($p = .018$), and in the reversed condition than the mistuned condition ($p = .005$), but there were no other significant differences: original-mistuned ($p = 1$); original-shuffled ($p = 1$); mistuned-shuffled ($p = 1$); reversed-shuffled ($p = .242$). In summary, participants recalled more artists than songs, but the artist recall ratio was highest in the reversed condition. We examine the implications of this and other results next.

Discussion

We confirmed that listeners can identify very brief song excerpts presented with their original acoustic

characteristics (Krumhansl, 2010) and in reverse (Schellenberg et al., 1999). We additionally discovered that listeners can identify mistuned plinks. However, the fact that listeners performed better in the original condition than in the mistuned condition suggests that listeners' utilize their implicit absolute pitch memories (Levitin, 1994; Schellenberg & Trehub, 2003) to identify plinks. Our most surprising novel finding was that listeners were able to identify temporally shuffled plinks for which dynamic spectral content was substantially disrupted while average spectral content was spared. Taken together, our observations of plink identification for mistuned, reversed, and shuffled plinks strongly suggest that neither canonical tuning or temporal order of acoustic events is necessary for plink identification. However, the fact that listeners identified the original plinks more accurately than the other conditions, responded to original plinks most quickly, and reported the highest recall confidence ratings for original plinks strongly indicates that plink perception benefits when stimulus acoustic characteristics, including absolute pitch and timbre content, match long-term memory. But, it is important to note that listener performance varied dramatically. Our best performing participant, who was in the original condition, recognized 74% of the plinks and recalled the song and artist names for 64% of the plinks. In comparison, our second best performing participant, also in the original condition, recognized 52% of the plinks and recalled the song and artist name for 32% of the plinks. All but 10 (who were all in the shuffled condition) of our 145 participants recognized at least one plink while all but 15 recalled at least one song or artist. Notably, the best performing participant was the only person to self-report the highest score on our absolute pitch questionnaire item, and this observation would be expected based on the notion that absolute pitch information facilitates plink perception.

Perhaps the most fascinating question raised by our results is, how can listeners identify mistuned, reversed, and shuffled plinks given that their acoustic content has never been encoded in long-term memory? Of course, humans routinely perform such auditory feats. For example, listeners can perceive a sentence spoken by talkers with different vocal ranges, speaking rates, and within different acoustic environments. Likewise, listeners can identify a given melody performed by singers with different ranges, performance tempi, and so on. Given that most listeners develop their implicit relative pitch perception abilities in infancy and retain them through adulthood (Trainor, 2005), it is not surprising that melodies remain recognizable when presented in

various keys. In general, shorter duration signals present less relative pitch information because melodic interval information builds over time. However, our 400 ms plinks are so brief that only a small number of rapidly produced events are present. If listeners do utilize relative pitch information to identify plinks, then relative pitch would contribute to plink recognition in both the original and mistuned conditions. Yet, relative pitch would probably not be as powerful a cue as timbre because timbre varies much more within and across plinks. Thus, relative pitch cues may contribute to plink identification, but they are not required for plink perception. What about absolute pitch cues? Schellenberg et al. (1999) found that listeners could match original, but not reversed (where absolute pitch cues are the same), 100 ms plinks to song names from a list. They interpreted this result to mean that "timbre is more important than absolute pitch for identifying popular recordings from very brief excerpts" (p. 645). However, this statement might be more appropriate for their stimuli, procedure, and results than ours because we did not observe any pairwise comparisons in which participants recognized or recalled reversed plinks any differently than mistuned plinks, although we did typically observe an advantage for mistuned and/or reversed plinks over shuffled plinks. The advantage for reversed plinks over shuffled plinks in our recognition and artist recall measures supports Schellenberg et al.'s assertion that listeners utilize local dynamic spectral cues. We extended their findings to show that listeners can use these dynamic cues even when they are presented in reverse (at least for our longer, 400 ms plinks). But, this finding does not imply that listeners do not use absolute pitch cues during plink perception. On the contrary, the fact that our participants identified original plinks more accurately than mistuned plinks supports the role of absolute pitch cues in plink perception. Just as pitch and timbre are entangled in the acoustic signal, it is likely true that pitch and timbre are entwined within listeners' musical memories. For example, Siedenburg and McAdams (2018) found that the accuracy of timbre ordering judgments in novel sound sequences decreases for sequences that vary in pitch, suggesting that pitch variability influences timbre memory.

Is plink perception a bottom-up process that depends exclusively on the signal or is it a top-down process that integrates context from long-term memory? There may be more than one perceptual-memory process at work (e.g., Kahneman, 2011; Wixted, 2007) and these processes may leverage multiple sources of information such as acoustic content within the signal or stored in long-term memory. Humans can rapidly process

acoustic information, even when that information is limited to brief durations, but one must not conflate plink duration with the time it takes to perceive a plink (Schellenberg et al., 1999). Identification might not occur until several seconds after plink presentation, and slower responses might be indicative of top-down processing. Recent research offers evidence for the bottom-up strategy. Jagiello et al. (2019) conducted EEG and pupillometry measurements during passive exposure to brief (750 ms) excerpts of familiar and unfamiliar music. They found that pupil dilation and EEG activity differed between familiar and unfamiliar excerpts within 400 ms of plink onset. We presented 400 ms plinks, and although recognition RT is not a direct measure of perceptual processes, our participants' RTs varied widely.

Given the combination of our data and our anecdotal observations, we cannot claim that plink identification depends exclusively on the acoustic content of each plink. Instead, we suggest a broader interpretation in which each plink activates a vast array of encoded acoustic content in which a detailed memory trace associated with song identity might ultimately emerge. On the one hand, plink identification could be considered an example of perceptual categorization. By analogy, rapid visual presentation is sufficient to trigger object categorization (VanRullen & Thorpe, 2001). On the other hand, plink identification could be considered an example of musical prediction in which the plink triggers slower perceptual or cognitive processes that enable a listener to imagine unheard events (Keller, 2012; Pearce & Wiggins, 2012; Rohrmeier & Koelsch, 2012). A similar argument for slower processing of plink stimuli was proposed by Bigand et al. (2005). Based on their finding that listeners could distinguish 250 ms plinks extracted from high- and low-moving music, they argued that the induced emotions "are too refined to be simply derived from basic emotional properties of sound . . . these responses required a cognitive appraisal" (p. 435). However, the fact that plink identification depends on a vast store of music memories does not necessarily mean that identification is a top-down process.

The critical question of whether plink perception is influenced by music memory is a single example of a broader, ongoing debate regarding whether cognitive processes can influence perception (Firestone & Scholl, 2016). In the introduction, we suggested that it may be possible for a listener to match a plink's transient acoustic events, whether in the original or reversed temporal order, to events stored in memory. Listeners could then utilize this low-level acoustic event information to build

a more complex song representation (Layman & Dowling, 2018). Alternatively, listeners might match a plink's average spectral content to music information in memory, but this latter process could proceed via a rapid, automatic bottom-up route or a slow, intentional, top-down route. Although we cannot decisively claim whether plink perception is bottom-up or top-down, we can comment on whether listeners utilize local and global information to identify plinks.

In their investigation of local and global feature processing in music perception, Bigand et al. (2009) demonstrated that listeners can distinguish between familiar and unfamiliar music excerpts (durations between 11–26 s) that have been acoustically scrambled (segment durations between 250–850 ms). In their discussion (p. 249), Bigand et al. cite another result from their lab in which listeners could distinguish familiar and unfamiliar music when the scrambled segment durations were as brief as 50 ms. Indeed, our results expand on Bigand et al.'s findings by showing that listeners can identify familiar plinks (400 ms duration) with scrambled segments (20 ms duration). We agree with Bigand et al. that the "color of sound" (p. 243), a timbral combination of voicing and instrumentation information, is perceptible in these signals. In the context of Bigand et al.'s scrambled excerpts that were approximately 50 times longer in duration than our plinks, it makes sense to consider each segment of the scrambled sequence as a local (e.g., 500 ms) representation of the global (e.g., 20 s) feature space. But in the context of our stimuli, each plink can be considered a local representation of the original, global (full-length) song. From that perspective, each 20 ms segment within a scrambled plink constitutes a hyperlocal representation of local feature space. The rapid presentation of randomly ordered segments in the shuffled plinks greatly diminishes a listener's ability to identify specific musical events that occur in either local or hyperlocal feature space. Thus, we propose that listeners can identify shuffled plinks, albeit at low baseline levels of performance, by comparing average spectral content to long-term memory. After all, a shuffled plink contains the same average spectral content as the original plink (Figure 1), though their hyperlocal spectral features differ. Additional support for this *average spectrum hypothesis* comes from the fact that listeners generally recalled artists more often than songs and that the artist recall ratio was highest in the reversed condition where listeners recalled artists nearly twice as often as songs. We expected these results based on the fact that a given song sounds more similar to songs performed by the same artist than songs performed by another artist. Alternatively, the artist identification

advantage might also occur because people can identify basic categories (i.e., artists) more readily than subordinate categories (i.e., songs; Rosch et al., 1976).

It is also possible that listeners can use both average and hyperlocal acoustic information to identify plinks, but additional research is needed to verify this. For example, if listeners can utilize hyperlocal features, then they probably will not need all 20 segments to identify a shuffled plink. Researchers could test this possibility by systematically reordering or removing segments (which will alter average spectral content) or by presenting listeners with a few hyperlocal acoustic events (event duration ≤ 20 ms). Here it is useful to note the variability in which listeners recognized and recalled plinks across our conditions (see Table 1). For example, 42% of listeners in the original condition and 22% of listeners in the shuffled condition recalled the song and artist names for Bodak Yellow (Cardi B). Yet, more listeners (72%) in the original condition and fewer (0%) in the shuffled condition recalled the song and artist names for Bad Blood (Taylor Swift). Patterns such as these may be due to random variability between listeners in each condition (e.g., listeners' familiarity with each song) or the degree to which the dynamic spectral content in the shuffled plinks serendipitously matched the original plinks. Although we randomized segments in the shuffled plinks, it is possible that a string of temporally adjacent segments from the original plinks also appeared in proximity to each other within the shuffled version, and this might be enough to trigger recall. However, we examined this possibility and determined that the number of serendipitously adjacent segments did not correlate with shuffled plink recognition (nor did several other measures of temporal orderliness; see Supplementary Materials accompanying the online version of this paper at mp.ucpress.edu). This result supports our claim that listeners recognize shuffled plinks by extracting their average spectra.

Although we considered creating a unique, randomly shuffled plink for each trial, we ultimately decided to use a single shuffled version of each plink. Thus, it is possible that our shuffle manipulation algorithm did not completely eliminate dynamic spectral cues. Although dividing each 400 ms plink into 20 randomly ordered segments substantially distorted the dynamic spectral content, the spectral content within each hyperlocal segment was maintained. Future researchers may want to investigate the degree to which dynamic spectral content is crucial for plink perception by gradually and systematically disrupting the dynamic spectral content of each plink. Investigations of alternative timbre smearing manipulations could also be fruitful. Researchers could explore the influence of segment

ordering, and potentially disconfirm the average spectrum hypothesis, by examining performance across several shuffled versions of any given original plink. Fine-grained acoustic comparisons of plinks across conditions could illuminate the importance of local and global spectral features in plink perception. And, although we believe that the non-zero recognition and recall proportions for the shuffled plinks provides strong evidence that listeners can use average spectral information to identify plinks, it is important to note that our results cannot provide proof because there are alternative explanations for listeners' abilities to identify them. For example, it is possible that some participants may have heard previous participants discuss the stimuli. Or, given the task demands, subjects may have inferred that the shuffled plinks comprised highly modified Pop songs. This might have been sufficient to activate memory representations of popular (familiar and common) music artists and songs, which may explain occasional successful responses by chance alone.

As we noted earlier, our results bolster the line of work that has shown that listeners encode absolute acoustic properties in long-term memory. One question that emerges from this is, how many exposures are necessary for listeners to form these absolute memories? Agus et al. (2010) showed that listeners can implicitly learn and recognize random noise patterns of 4 s duration in as little as four presentations. Schellenberg et al. (2019) exposed participants to novel melodies in a lab setting. After a brief, ten-minute delay, they presented these melodies again, along with new melodies. Critically, they manipulated the pitch (Experiment 1) and tempo (Experiment 2) of the learned melodies to determine how these acoustic factors influence recall. Schellenberg et al. found that changes to key, regardless of magnitude, had a negative effect on melody recognition, as did tempo, though the effect was weak. Taken together, these results demonstrate that listeners encode pitch and tempo information within the first exposure to a piece of music, and that this information can be accessed for future tasks such as song recognition. By extension, we suspect that listeners could identify plinks from songs they have only heard and attended to once, but this claim remains to be tested in the laboratory.

An examination of the recognition and recall proportions across plinks (Table 1) and conditions (Table 3) might lead the reader to wonder how our observations compare to prior work. Participants in the original condition recalled the song and artist names for 15% of the plinks. In comparison, Krumhansl (2010, p. 343) and Thiesen et al. (2020, p. 341) reported identification rates of 26% and 7% for their 400 ms plinks. The variation in

plink perception across studies suggests that stimulus and sample selection can have a substantial effect on the results. We chose songs that we expected to be maximally familiar to our sample whereas Thiesen et al. (2020) chose a mixture of recent and older songs based on stimuli from Krumhansl's (2010) work. The distribution of recall rates across studies indicates that plink identification estimates are sensitive to participants' familiarity with the songs used to create the plink stimuli. Relatedly, Spivack et al. (2019) presented listeners with song stimuli from top-charting popular music between 1940 to 2015. Their participants, who were similarly aged as ours, reported highest recognition for songs released between 2001 to 2015 and progressively lower recognition for songs released before 2000. Thus, considering that listeners are most familiar with the music they typically encounter, plink researchers should expect recall performance rates for any given stimulus set to decrease over time. This has consequences for future efforts to systematically replicate plink perception effects: the songs that are familiar to participants today are not the same songs that will be familiar to participants tomorrow (Greenfield, 2017). Instead of using the same stimuli as prior work, researchers may want to consider implementing a systematic stimulus selection rule such as the one employed by Spivack et al. In that case, some stimuli will be the same as prior work but recently released, chart-topping songs would be added for each new investigation.

There are few important limitations of our method and results. First, our findings should not be taken to indicate that some songs in our stimulus set are objectively more recognizable than others, although that may be true. As Krumhansl (2010) and Thiessen et al. (2019) pointed out, the place in the song from which a plink is extracted can have substantial effects on plink identification. We agree, and, although we do not know how plink perception would have been affected if we had extracted plinks from different locations within the song choruses, we believe that our general pattern of results would hold true. Another limitation of our work is that, unlike many RT experiments, we did not strongly emphasize speeded responses in our task instructions. The fact that we did not identify many RT differences, other than fastest responses for original plinks, could indicate that participants may have prioritized response accuracy over speed. It could also indicate that our RT measure was not as sensitive as it could have been (if, for example, we measured speeded response after a single plink presentation). Researchers who intend to investigate the time course by which listeners access implicit absolute pitch and timbre memories should be careful

to design their studies to maximize RT validity and reliability. Furthermore, researchers should try to distinguish RT associated with recognition and recall. Additionally, we recommend that researchers who measure plink recall should carefully curate their stimulus list to eliminate songs with alternative titles, featured artists, and other characteristics that could complicate response coding. Participant factors might also have influenced performance. Although we selected our stimuli to be largely recognizable to the sample we recruited, it is likely that differences in participants' musical tastes and listening habits caused variation in stimulus familiarity that our between-subjects design could not account for. Researchers could utilize within-subjects designs to investigate the role of participant familiarity on plink recognition and recall. Finally, we have no idea how participants would have performed if they were in a different experimental context that did not explicitly instruct them to recall song and artist names. It is reasonable to assume that some degree of recognition and recall accuracy is explained by spreading activation and priming (e.g., Bharucha, 1987). For example, it may be the case that participants who successfully recognize one plink will be more likely to recognize another due to remembered associations between songs and artists. Unfortunately, our stimulus set may have amplified this possibility because we included several songs performed by a single artist. However, it is also conceivable that such associations may have produced interference that ultimately reduced performance. To examine the possibility of plink priming effects, researchers could embed musical plinks with many other natural and artificial sounds or compare performance patterns between groups in which single or multiple songs from a given artist are included. Finally, though our results are compatible with the notion that relative pitch contributes to plink identification, our design was incapable of establishing such an effect. We recommend that researchers investigate the possible role of relative pitch cues by carefully manipulating pitch intervals within plinks. We suspect that relative pitch manipulations would reduce plink identification rates and thereby support the role of relative pitch in plink perception.

The plink paradigm emerged when researchers began to ask how quickly humans can identify music genres (Gjerdingen & Perrott, 2008) and which acoustic characteristics enable humans to recognize songs (Schellenberg et al., 1999). Despite the specificity of these original research questions, ongoing research has revealed that the plink paradigm is relevant to broad theoretical investigations about the rapidity of perceptual

processes, the overlap between perception and cognition, the basic structure of the mind, and computational models of human performance. In a footnote, Krumhansl (2010) stated that Shazam (Apple, Inc., 2020), a popular music recognition and discovery application, could not identify her plinks. That is also true for our set of original, mistuned, reversed, and shuffled plinks. However, Shazam can reliably identify original music given a few seconds of input, and we also found that it could identify longer, mistuned excerpts. However, unlike human listeners, Shazam does not identify these longer excerpts when they are reversed or temporally shuffled (at least with our shuffling parameters). The differences in music identification performance between humans and computers underscore the different strategies that each employs. For example, Shazam appears to need more than one second of input and it relies on temporally anchored acoustic characteristics (Wang, 2003, 2006) such that major disruptions in the time domain decrease identification performance. Our results show that human listeners, on the other hand, are more flexible in their auditory search strategies. Of course, computer algorithms can be reprogrammed,

and we hope that our results inspire the development of more robust auditory search engines. Humans have remarkable auditory perception abilities in part because they maintain vast stores of relative and absolute acoustic content in memory. In conclusion, we have shown that listeners utilize absolute acoustic cues to identify very brief musical sounds and that they utilize dynamic and average spectral content to perceive plinks. Given these observations, future researchers should be better positioned to investigate which sources of acoustic information are necessary or sufficient for plink perception.

Author Note

Stimuli, data files, and other materials are available at <https://osf.io/ma589/>. We gratefully acknowledge Hannah Madden and Eduardo Elias for their assistance with data collection. We thank two anonymous reviewers for their feedback.

Correspondence concerning this article should be addressed to James T. Mantell, St. Mary's College of Maryland, 47645 College Dr., St. Mary's City, MD 20686. E-mail: jtmantell@smcm.edu

References

- AGUS, T. R., SUIED, C., THORPE, S. J., & PRESSNITZER, D. (2012). Fast recognition of musical sounds based on timbre. *Journal of the Acoustical Society of America*, 131(5), 4124–4133. <https://doi.org/10.1121/1.3701865>
- AGUS, T. R., THORPE, S. J., & PRESSNITZER, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron*, 66(4), 610–618. <https://doi.org/10.1016/j.neuron.2010.04.014>
- ALEXANDER, J. M., JENISON, R. L., & KLUENDER, K. R. (2011). Real-time contrast enhancement to improve speech recognition. *PLOS ONE*, 6(9), e24630. <https://doi.org/10.1371/journal.pone.0024630>
- APPLE, INC. (2020). *Shazam* (Version 11.3.0) [Mobile app]. Google Play. <https://play.google.com/store/apps/details?id=com.shazam.android>
- AUDACITY TEAM. (2018). *Audacity* (Version 2.2) [Computer software]. <https://www.audacityteam.org/>
- BELFI, A. M., KASDAN, A., ROWLAND, J., VESSEL, E. A., STARR, G. G., & POEPPPEL, D. (2018). Rapid timing of musical aesthetic judgments. *Journal of Experimental Psychology: General*, 147(10), 1531–1543. <http://dx.doi.org/10.1037/xge0000474>
- BEN-HAIM, M. S., EITAN, Z., & CHAJUT, E. (2014). Pitch memory and exposure effects. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 24–32. <https://doi.org/10.1037/a0033583>
- BHARUCHA, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, 5(1), 1–30. <https://doi.org/10.2307/40285384>
- BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2), 115–147. <https://doi.org/10.1196/annals.1360.036>
- BIGAND, E., DELBÉ, C., GÉRARD, Y., & TILLMANN, B. (2011). Categorization of extremely brief auditory stimuli: Domain-specific or domain-general processes? *PLOS ONE*, 6(10), e27024. <https://doi.org/10.1371/journal.pone.0027024>
- BIGAND, E., FILIPIC, S., & LALITTE, P. (2005). The time course of emotional responses to music. *Annals of the New York Academy of Sciences*, 1060(1), 429–437. <https://doi.org/10.1196/annals.1360.036>
- BIGAND, E., GÉRARD, Y., & MOLIN, P. (2009). The contribution of local features to familiarity judgments in music. *Annals of the New York Academy of Sciences*, 1169(1), 234–244. <https://doi.org/10.1111/j.1749-6632.2009.04552.x>
- BOERSMA, P., & WEENINK, D. (2018). *Praat: Doing phonetics by computer* (Version 6.0) [Computer program]. <http://www.praat.org/>
- BOOTHROYD, A., MULHEARN, B., GONG, J., & OSTROFF, J. (1996). Effects of spectral smearing on phoneme and word recognition. *Journal of the Acoustical Society of America*, 100(3), 1807–1818. <https://doi.org/10.1121/1.416000>

- COHEN, B. H. (2013). *Explaining psychological statistics* (4th ed.). John Wiley & Sons.
- FILIPIC, S., TILLMANN, B., & BIGAND, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin and Review*, 17(3), 335–341. <https://doi.org/10.3758/PBR.17.3.335>
- FIRESTONE, C., & SCHOLL, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, e229. <https://doi.org/10.1017/S0140525X15000965>
- GJERDINGEN, R. O., & PERROTT, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research*, 37(2), 93–100. <https://doi.org/10.1080/09298210802479268>
- GREENFIELD, P. M. (2017). Cultural change over time: Why replicability should not be the gold standard in psychological science. *Perspectives on Psychological Science*, 12(5), 762–771. <https://doi.org/10.1177/1745691617707314>
- HINTZMAN, D. L., CAULTON, D. A., & LEVITIN, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory and Cognition*, 26(3), 449–462. <https://doi.org/10.3758/BF03201155>
- HOU, Z., & PAVLOVIC, C. V. (1994). Effects of temporal smearing on temporal resolution, frequency selectivity, and speech intelligibility. *Journal of the Acoustical Society of America*, 96(3), 1325–1340. <https://doi.org/10.1121/1.410279>
- ISNARD, V., TAFFOU, M., VIAUD-DELMON, I., & SUIED, C. (2016). Auditory sketches: Very sparse representations of sounds are still recognizable. *PLOS ONE*, 11(3), e0150313. <https://doi.org/10.1371/journal.pone.0150313>
- JAGIELLO, R., POMPER, U., YONEYA, M., ZHAO, S., & CHAIT, M. (2019). Rapid brain responses to familiar vs. unfamiliar music: EEG and pupillometry study. *Scientific Reports*, 9(1), 1–13. <https://doi.org/10.1038/s41598-019-51759-9>
- KAHNEMAN, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- KELLER, P. E. (2012). Mental imagery in music performance: Underlying mechanisms and potential benefits. *Annals of the New York Academy of Sciences*, 1252(1), 206–213. <https://doi.org/10.1111/j.1749-6632.2011.06439.x>
- KIM, Y. E., WILLIAMSON, D. S., & PILLI, S. (2006). Towards quantifying the “album effect” in artist identification. *Proceedings of The International Society of Music Information Retrieval, Canada*, 393–394. <http://doi.org/10.5281/zenodo.1415722>
- KRUMHANSL, C. L. (2010). Plink: “Thin slices” of music. *Music Perception*, 27(5), 337–354. <https://doi.org/10.1525/mp.2010.27.5.337>
- KRUMHANSL, C. L., & ZUPNICK, J. A. (2013). Cascading reminiscence bumps in popular music. *Psychological Science*, 24(10), 2057–2068. <https://doi.org/10.1177/0956797613486486>
- LAYMAN, S. L., & DOWLING, W. J. (2018). Did you hear the vocalist? Differences in processing between short segments of familiar and unfamiliar music. *Music Perception*, 35(5), 607–621. <https://doi.org/10.1525/mp.2018.35.5.607>
- LEVITIN, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics*, 56(4), 414–423. <https://doi.org/10.3758/BF03206733>
- LEVITIN, D. J., & COOK, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception and Psychophysics*, 58(6), 927–935. <https://doi.org/10.3758/bf03205494>
- LEVITIN, D. J., & MENON, V. (2005). The neural locus of temporal structure and expectancies in music: Evidence from functional neuroimaging at 3 Tesla. *Music Perception*, 22(3), 563–575. <http://www.jstor.org/stable/10.1525/mp.2005.22.3.563>
- LEVITIN, D. J., & ROGERS, S. E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences*, 9(1), 26–33. <https://doi.org/10.1016/j.tics.2004.11.007>
- MACE, S. T., WAGONER, C. L., TEACHOUT, D. J., & HODGES, D. A. (2011). Genre identification of very brief musical excerpts. *Psychology of Music*, 40(1), 112–128. <https://doi.org/10.1177/0305735610391347>
- MANDLER, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of recognition. *Perspectives on Psychological Science*, 3(5), 390–399. <https://doi.org/10.1111/j.1745-6924.2008.00087.x>
- McKELLAR, J. L., & COHEN, A. (2015). Identification of thin slices of music by university students in PEI. *Canadian Acoustics*, 43(3), 88–89. <https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2810>
- NORDSTRÖM, H., & LAUKKA, P. (2019). The time course of emotion recognition in speech and music. *Journal of the Acoustical Society of America*, 145(5), 3058–3074. <https://doi.org/10.1121/1.5108601>
- PEARCE, M. T., & WIGGINS, G. A. (2012). Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science*, 4(4), 625–652. <https://doi.org/10.1111/j.1756-8765.2012.01214.x>
- PEIRCE, J. W., GRAY, J. R., SIMPSON, S., MACASKILL, M. R., HÖCHENBERGER, R., SOGO, H., ET AL. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51, 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- PLAZAK, J., & HURON, D. (2011). The first three seconds: Listener knowledge gained from brief musical excerpts. *Musicae Scientiae*, 15(1), 29–44. <https://doi.org/10.1177/1029864910391455>

- ROHRMEIER, M. A., & KOELSCH, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology*, 83(2), 164–175. <https://doi.org/10.1016/j.ijpsycho.2011.12.010>
- ROSCH, E., MERVIS, C. B., GRAY, W. D., JOHNSON, D. M., & BOYES-BRAEM, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382–439. [https://doi.org/10.1016/0010-0285\(76\)90013-X](https://doi.org/10.1016/0010-0285(76)90013-X)
- SCHELLENBERG, E. G., IVERSON, P., & MCKINNON, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin and Review*, 6(4), 641–646. <https://doi.org/10.3758/BF03212973>
- SCHELLENBERG, E. G., & TREHUB, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, 14(3), 262–266. <https://doi.org/10.1111/1467-9280.03432>
- SCHELLENBERG, E. G., WEISS, M. W., PENG, C., & ALAM, S. (2019). Fine-grained implicit memory for key and tempo. *Music and Science*, 2, 1–14. <https://doi.org/10.1177/2059204319857198>
- SCHULKIND, M. D. (2004). Serial processing in melody identification and the organization of musical semantic memory. *Perception and Psychophysics*, 66(8), 1351–1362. <https://doi.org/10.3758/BF03195003>
- SCHULZE, K., JAY DOWLING, W., & TILLMANN, B. (2011). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception*, 29(3), 255–267. <https://doi.org/10.1525/mp.2012.29.3.255>
- SCHWEINBERGER, S. R., HERHOLZ, A., & SOMMER, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language, and Hearing Research*, 40(2), 453–463. <https://doi.org/10.1044/jslhr.4002.453>
- SIEDENBURG, K., & McADAMS, S. (2018). Short-term recognition of timbre sequences: Music training, pitch variability, and timbral similarity. *Music Perception*, 36(1), 24–39. <https://doi.org/10.1525/mp.2018.36.1.24>
- SIEDENBURG, K., & MÜLLENSIEFEN, D. (2017). Modeling timbre similarity of short music clips. *Frontiers in Psychology*, 8, 639. <https://doi.org/10.3389/fpsyg.2017.00639>
- SPIVACK, S., PHILIBOTTE, S. J., SPILKA, N. H., PASSMAN, I. J., & WALLISCH, P. (2019). Who remembers the Beatles? The collective memory for popular music. *PLOS ONE*, 14(2), e0210066. <https://doi.org/10.1371/journal.pone.0210066>
- SUIED, C., AGUS, T. R., THORPE, S. J., MESGARANI, N., & PRESSNITZER, D. (2014). Auditory gist: Recognition of very short sounds from timbre cues. *Journal of the Acoustical Society of America*, 135(3), 1380–1391. <http://dx.doi.org/10.1121/1.4863659>
- THIESEN, F. C., KOPIEZ, R., MÜLLENSIEFEN, D., REUTER, C., & CZEDIK-EYSENBERG, I. (2020). Duration, song section, entropy: Suggestions for a model of rapid music recognition processes. *Journal of New Music Research*, 49(4), 334–348. <https://doi.org/10.1080/09298215.2020.1784955>
- THIESEN, F. C., KOPIEZ, R., REUTER, C., & CZEDIK-EYSENBERG, I. (2019). A snippet in a snippet: Development of the Matryoshka principle for the construction of very short musical stimuli (plinks). *Musicae Scientiae*, 24(4), 515–529. <https://journals.sagepub.com/doi/full/10.1177/1029864918820212>
- TRAINOR, L. J. (2005). Are there critical periods for musical development? *Developmental Psychobiology*, 46(3), 262–278. <https://doi.org/10.1002/dev.20059>
- VANRULLEN, R., & THORPE, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, 13(4), 454–461. <https://doi.org/10.1162/08989290152001880>
- WANG, A. (2003). An industrial-strength audio search algorithm. *Proceedings of the International Conference on Music Information Retrieval*. <https://doi.org/10.5281/zenodo.1416340>
- WANG, A. (2006). The Shazam music recognition service. *Communications of the ACM*, 49(8), 44–48. <https://doi.org/10.1145/1145287.1145312>
- WHITE, B. (1960). Recognition of distorted melodies. *The American Journal of Psychology*, 73(1), 100–107. <https://doi.org/10.2307/1419120>
- WIXTED, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114(1), 152–176. <https://doi.org/10.1037/0033-295X.114.1.152>