Salience of Frequency Micro-modulations in Popular Music

Michel Bürgel & Kai Siedenburg
University of Oldenburg, Oldenburg, Germany

Singing voices attract auditory attention in music unlike other sounds. In a previous study, we investigated the salience of instruments and vocals using a detection task in which cued target sounds were to be detected in musical mixtures. The presentation order of cue and mixture signals influenced the detection of all targets except the lead vocals, indicating that listeners focus on voices regardless of whether these are cued or not, highlighting a unique vocal salience in music mixtures. The aim of the present online study was to investigate the extent to which phonological cues, musical features of the main melody, or frequency micro-modulation (FMM) inherent in singing voices contribute to this vocal salience. FMM was either eliminated by using an autotune effect (Experiment 1) or transferred to other instruments (Experiment 2). Detection accuracy was influenced by presentation order for all instrumental targets and the autotuned vocals, but not for the unmodified vocals, suggesting that neither the phonological cues that could provide a facilitated processing of speech-like sounds nor the musical features of the main melody are sufficient to drive vocal salience. Transferring FMM from vocals to instruments or autotuned vocals reduced the magnitude of the order effect considerably. These findings suggest that FMM is an important acoustical feature contributing to vocal salience in musical mixtures.

Received: July 19, 2022, accepted July 15, 2023.

Key words: auditory attention, autotune, frequency micro-modulations, vocal salience, timbre

Who has not experienced it: While listening to music, the ear seamlessly picks up a catchy vocal melody from a musical mix. A melody emerges in the mind of the listeners, seemingly independent from the musical background that it was embedded in. Notwithstanding the ease of auditory processing, multi-instrumental music confronts listeners with complex acoustic scenes, in which instruments and voices overlap in both time and frequency. Despite the potential complexity of musical scenes, the auditory system analyzes and groups musical mixtures into representations of individual streams. This ability to organize sounds into perceptual streams is referred to as auditory scene analysis (ASA; Bregman & McAdams, 1994). This framework assumes that ASA is determined by primitive (bottom-up) and schema-driven (top-down) processing. The latter is thought to incorporate processes of scene parsing based on attention, memory, and knowledge. Selective attention in ASA has been studied using an interleaved melody recognition paradigm with simple melodies (Bey & McAdams, 2002), which has listeners detect a target sound in a mixture. The target can be presented before or after the mixture and the resulting difference in detection accuracy is assumed to be due to processes of selective attention. In a previous study (Bürgel et al., 2021), we found that all sound categories except the lead vocals showed effects of selective attention. Because accuracy was particularly high and independent of selective attention for vocals, we dubbed this pattern of results *vocal salience*. Here, we wished to further explore the basis of vocal salience in popular music. Generally, this approach extends previous research by using mixtures of popular music as highly realistic and representative stimuli for ASA research.

Auditory attention, such as the reflex-like focusing on a loud sound or deliberate listening to an instrument in a mixture, modulates the cognitive representation of the acoustic scene by allocating processing resources to distinct elements of a scene (e.g., Shamma et al., 2011; Sussman, 2017). Studies of auditory attention in musical scenes found that the voice occupies a unique role among other sound sources, enabling the voice to stand out from other instruments in a mixture: When human listeners are asked to recognize isolated voices and instruments, responses to voices occur faster and with higher accuracy (Agus et al., 2012). Moreover, voice sounds require a shorter time of exposure for recognition compared to other musical instrument sounds (Isnard et al., 2019; Suied et al., 2014). When comparing vocal melodies and instrumental melodies, previously presented vocal melodies are more precisely recognized compared to instrumental melodies (Weiss et al., 2012). Neurophysiological experiments underpin this unique role of the vocals, showing an enhanced cortical

Music Perception, volume 41, issue 1, pp. 1–14, issn 0730-7829, electronic issn 1533-8312. DOI: https://doi.org/10.1525/mp.2023.41.1.1

response when vocal signals are presented in isolation among speech and non-vocal environmental sounds (Belin et al., 2022; Belin et al., 2000), and among other instruments (Gunji et al., 2003; Levy et al., 2001). Further, when presented in a musical mixture, specific neural populations were found that respond distinctively to music with singing voices but not to instrumental music (Norman-Haignere et al., 2022).

This facilitated processing of vocals also plays out in multi-instrumental musical mixtures. Previously, we investigated the detection of cued target instruments and voices in short excerpts of popular music mixtures (Bürgel et al., 2021). The cue consisted of an isolated instrument or voice and was either presented before or after the mixture. Notably, all target signals except the lead vocals showed a clear surplus of detection accuracy when the target cue was presented before the mixture, highlighting the intrinsic salience of the vocals that attracts the listener's attention regardless of the presentation of a cue. This salience persisted and was unmatched by other instruments, even when the instruments and vocals were matched in sound level or were spectrally filtered to pass through the mixture unmasked.

The question arises as to which features of vocal signals contribute to their unique role among natural sounds. Here, we considered three candidate features. First, it may seem reasonable to suggest that the unique salience of vocals could arise from the phonological information they contain. Language specific processing may potentially activate increased attentional resources (Signoret et al., 2011). Second, another feature contributing to the unique presence of the vocals could be their favorable musical role in the multi-instrument mixtures. In Western popular music, the lead vocals contribute the main melody of a song and thus are composed to possess a prominent role with respect to the accompanying instruments and background vocals. When listening to music hierarchically structured into main melody and accompaniment, previous studies have shown that attention is drawn towards the main melody (Ragert et al., 2014).

Third, a more acoustically based candidate feature may be related to frequency micro-modulation (FMM). Here, we understand FMM as non-stationary frequency changes in acoustic signals, usually less than one semitone, which are not perceived as irregular or as intonation errors. In singing, FMM tends to be caused by imperfect control of intonation caused by vocal-motor control adjustments of the human voice (Hutchins et al., 2014) and is present even in highly trained singers (e.g., Hutchins & Campbell, 2009; Mori et al., 2004; Sundberg et al., 1996). Even though pitch detection for vocals seems to be less precise than for musical instruments (Gao & Oxenham, 2022; Hutchins et al., 2012; Sundberg et al., 2013), FMM influences the perception of intonation (Larrouy-Maestri & Pfordresher 2018), is known to facilitate the prominence of vowel sounds (Marin & McAdams, 1991; McAdams, 1989), and evokes cortical responses that can be traced by neurophysiological measurements (Saitou et al., 2005). Experiments with speech signals indicate that both the exaggeration and reduction of the modulations result in decreased speech intelligibility (Miller et al., 2010); frequency modulations naturally inherent in speech signals were associated with highest speech intelligibility scores.

The purpose of the present study is to further investigate the unique ability of the vocals to be the focal point of auditory attention in musical scenes (vocal salience), which was found in our previous experiments (Bürgel et al., 2021). More precisely, we investigate how these three candidate features contribute to vocal salience. We analyze the role of FMM as well as phonological cues in natural singing voices, either by eliminating the modulations in the vocals (Experiment 1) or by transferring the modulations to instruments (Experiment 2). We further examine how having instruments play the vocal melody affects their salience in the mixture (Experiment 1 & Experiment 2). We use the same experimental paradigm as in our prior experiments (Bürgel et al., 2021): participants were asked to detect a cued target signal (vocal or instrument) embedded in a mixture of multiple instruments. Because detection accuracy is influenced not only by the salience of the target but also by factors such as sound level or spectral masking (Bürgel et al, 2021; Siedenburg et al., 2019), we test the effect of the presentation order of target cue and mixture to isolate how the detection of the target signal is modulated by auditory attention. For one half of the participants, the target cue is presented first and followed by the mixture, allowing the cue to be used to "search" the mixture for the target. This order is used to measure detection accuracy in a facilitated listening situation where participants have prior knowledge of the target. For the other half of participants, the presentation order is reversed, with the mixture presented first, so that the detection of targets strongly depends on the salience of the target in the mixture. A comparison between both presentation orders allows us to quantify the influence of the effect of selective attention through the surplus of the accuracy in the target-mixture condition compared to the mixture-target condition.

For the conditions where FMM is eliminated from the vocal signals, we speculate on two possible outcomes:

Either the facilitated detection for singing voices remains intact because it is driven by phonological cues that encourage a facilitated processing of speech-like sounds, and that are retained throughout the pitch quantization. Alternatively, detection of singing voices degrades, because vocal salience is a result of the human sensitivity towards FMM. Considering the role of the melodic material, we speculate that in trials in which instruments replace the vocals and play the main melody detection accuracy is clearly facilitated. For transferring the melody and FMM of the lead vocals to instruments, we expect that the presence of FMM that are uncommon for the instruments introduces a cue that results in an increase of detection accuracy compared to conditions presenting the main melody without FMM. If the FMM is driving the facilitated detection of vocals, this transfer of FMM may decrease or even eliminate the effect of presentation order.

Method

PARTICIPANTS

All participants were recruited via an online call for participation on the e-learning platform of the University of Oldenburg. The call included a briefing, a link to the online experiments, and inclusion criteria such as the use of headphones, a stable internet connection, and self-reported normal hearing. Participants could take part in the experiment online at any time during a one-month time window. Participants who took part in Experiment 1 were not permitted to take part in Experiment 2. A total of 69 participants (age: M = 25.1, SD = 3.5) took part in Experiment 1 and 70 participants (age: M = 24.7, SD = 3.5) in Experiment 2.

In Experiment 1, the overall scores of individual listeners were distributed bimodally, with three participants exhibiting drastically worse results (< 60% correct responses) compared to most other listeners, indicating that they did not actively participate in the experiment and were therefore discarded from the analysis. A histogram with overall accuracies of included and excluded participants is shown in the Supplementary Materials (see Individual results) accompanying the online version of this paper at mp.ucpress.edu. The same was true for two participants in Experiment 2 (< 60% correct responses). The results of 67 participants (age: M = 25.1, SD = 3.2) in Experiment 1 and 67 participants (age: M = 24.7, SD = 3.4) in Experiment 2 were analyzed. In both experiments, participants were randomly assigned to one of two groups that determined the order in which the target cue and mixture were presented: 33 participants (age: M = 24.8, SD = 3.3) in Experiment 1 and 33 participants (age: M = 24.8, SD = 3.3) in Experiment 2 were assigned to the order in which the target was presented before the mixture. For the reverse order, 34 (age: M = 25.3, SD = 3) participants in Experiment 1 and 35 participants (age: M = 24.7, SD = 3.8) in Experiment 2 were assigned. We acquired information on the participants' musical abilities using a subset of the Gold-MSI (Müllensiefen et al., 2014) consisting of nine questions on music perception abilities and seven questions on music training.

STIMULI AND TASK

Stimuli were generated in MATLAB by extracting two-second excerpts of a single target instrument or vocals and a mixture of multiple instruments and vocals from a multitrack music database ("MedleyDB," https://medleydb.weebly.com/), see Figure 1A for a schematic. The database consisted of 127 royalty-free songs covering a wide range of popular music genres, with individual audio files for each instrument and vocals. The majority of the songs had English lyrics. Instruments and vocals were mixed so that the overall mix adhered to the conventions of popular music. We coarsely categorized the instruments and vocals in the database as: Backing Vocals, Bass, Drums, Guitars, Lead Vocals, Piano, Percussion, Strings, Synthesizer, Winds. For each excerpt, a to-be attended instrument or vocal was chosen (target). Remaining instruments or voices in the excerpts that did not belong to the same category as the target functioned as maskers (mixture). Instruments or voices in the excerpt that belonged to the same category as the target were not included in the excerpt. In the case where the lead vocals were assigned as the target, all backing vocals were also excluded. Guitar, synthesizer, and winds were selected as instrument targets and the category lead vocals was selected as vocal targets. For guitar, synthesizer and wind targets that were adapted to the main melody, excerpts of the lead vocals were used as the basis.

To examine song excerpts for potential stimuli, we computed an instrument and vocal activity analysis for each song, indicating which instrument or vocals were likely audible in a given time frame. The activity analysis was created by calculating the sound level of each instrument and vocal in each song using a 500 ms sliding window. In each window, the root-mean-square value (RMS) of the sound level was calculated. For each instrument or vocal, the instrument or vocal was considered active in a time window, when the sound level in the window was above -20 dB relative to the maximum sound level of the entire song of the respective instrument or vocal. To further control the complexity of

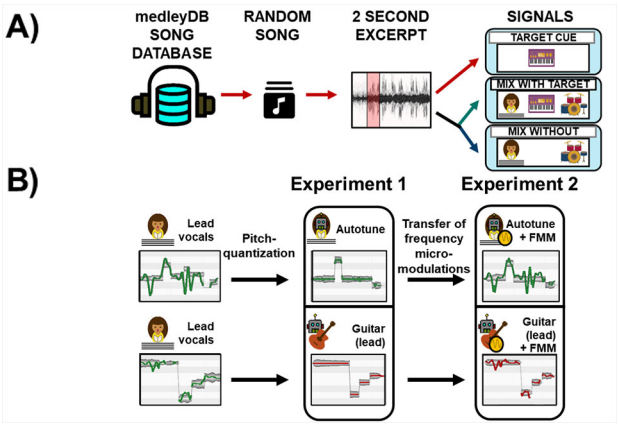

FIGURE 1. Schematic overview of the methods. (A) Stimulus extraction: Short excerpts from the open source "medleyDB" multitrack database were used. Songs were drawn randomly without replacement. From each excerpt, two signals were extracted: One signal containing only the target signal, another signal either containing the mixture with or without the target signal. See the text for details. (B) Vocal manipulation: Lead vocal excerpts were pitch-quantized to create autotune or instrument main melody targets (lead) in Experiment 1. The original frequency trajectory of the unquantized vocal tracks was reapplied to the autotune und instrument main melody targets in Experiment 2. The gray waveform represents the amplitude of the excerpt over time. Within the waveform, colored lines indicate the frequency trajectory. The light and dark gray shades indicate divisions of a chromatic scale in semitone steps.

musical scenes in our stimuli, we removed all time windows from the activity map in which fewer than five and more than nine instruments or vocals were active. For each target category, we drew a 2000 ms excerpt with four adjacent, previously unused 500 ms time windows in which the target category and up to seven other vocals or instrument categories were considered active. Time slices were drawn from pseudo-randomly selected songs, with a preference to use the same song as infrequently as possible. In this way a total of 30 excerpts for each instrument target and 150 excerpts for vocal targets were drawn. The excerpts for the vocal target were then subdivided to be used either as vocal target, pitch-quantized vocal targets (autotune), or instrument targets playing the main melody. Furthermore, the excerpts contained sung English words, which could foster a potential facilitated processing of phonological features.

One hundred and twenty vocal tracks were pitch quantized using the pitch correction software *Melodyne* (Melodyne Version 5, Celemony Software). The corresponding manipulation of FMM is illustrated in Figure 1B by two exemplary excerpts and in the Supplementary Materials Figure 2. Quantization was set to both match pitch to a tempered scale and to eliminate all FMM, resulting in a robotic voice quality typical of the autotune effect. Thirty vocal tracks were modified in this way and were used as targets for the "autotune" category. The pitch of the remaining 90 quantized vocal tracks was used as a basis for the instrument main melody targets by having the melodies being played by three different MIDI-based instruments that corresponded to a guitar, synthesizer, or wind sound, thus creating 30 tracks for each of the three instruments. MIDI notes were programmed manually to accurately match the vocals in pitch, on- and offset times. For Experiment 2, the original frequency trajectory of the unquantized vocal tracks was reapplied to the autotuned vocals und instrument main melody targets by using the *Auto-Tune Pro* Plugin (Auto-Tune Pro, Antares).

Three monophonic signals were compiled from each two-second excerpt: 1) a signal containing only the target, 2) a signal containing a mixture of five to eight instruments or vocals from non-target categories plus the target, and 3) a signal containing a mixture of six to nine instruments or vocals without the target. For mixtures, the full number of instruments that were also present in the original excerpt of the song were used. A logarithmic fade-in and fade-out with a duration of 200 ms was applied to the beginning and end of all extracted signals. The sound level ratio between the target and the mixture was adjusted to -10 dB (cf., Bürgel et al., 2021). For half of the trials, the mixture signals were arranged to contain the target signals; for the other half, the mixture did not contain the target signal. To prevent the presence of the three MIDI instruments from serving as a cue of the target, the lead vocals of the mix were replaced by one of the MIDI instruments using the same sound level as the vocals in one-third of the excerpts where an accompanying instrument was the target. Stimuli were created using the isolated target signal and a mixture signal in which the target was either present or absent. A 500 ms pause was inserted between the two signals, resulting in a total stimulus duration of 4500 ms. By interchanging the presentation order of target and mixture signal, two order conditions were created: In the "Target-Mixture" condition, the target signal was followed by a pause and the mixture signal; in the "Mixture-Target" condition, the presentation order was reversed. For use on the online platform, stimuli were converted from WAV format to MP3 at a bit rate of 320 kbit/s. Example stimuli and sound samples are provided on the website: https://uol.de/en/musik-wahrnehmung/sound-examples/akrs

PROCEDURE

The experiments were approved by the ethics committee of the University of Oldenburg and conducted online via the web platform www.testable.org. Experiment 1 and Experiment 2 were identical in design, used the same song excerpts, and differed only in the absence (Experiment 1) or presence (Experiment 2) of FMM in the autotune and main melody instrument targets. Participants were automatically assigned to one of two groups, determining the presentation order of target cue and mixture. In the first group all stimuli appeared in the "Target-Mixture" presentation order, whereas for the second group the order was reversed to the "Mixture-Target" order. Each experiment used the same excerpts and was structured into five consecutive segments.

In the first segment, participants had to complete a headphone screening task based on Milne et al. (2021). Here, a sequence of three white noise signals were presented, with one of the noise signals being phase shifted by 180 degrees in a narrow frequency band at around 600 Hz on the left headphone channel. When headphones were worn, the phase shift is perceived as a narrow tone embedded in the broadband noise. The task began with an instruction and a presentation of the noise signal, a 600 Hz tone in isolation and three mixtures of the tone in noise. Listeners had to detect the tone and passed the test if five out of six responses were correct. Participants who did not pass the headphone screening were returned to the instruction panel and reminded that they must pass the headphone screening before they were allowed to continue.

After the headphone screening, three song excerpts were presented to provide an impression of the dynamic range of the stimuli. During the presentation, participants were instructed to adjust the sound to a comfortable level. This was followed by a training phase, to familiarize participants with the detection task. Participants were presented with stimuli that were very similar but different from those used in the main experiment and were asked whether the target was present or absent in the mixture. Participants were allowed as much time as they needed to respond to the questions. To help participants understand the task, feedback was given after each answer. One stimulus with target and one without target in the mixture among the target categories lead vocals, autotune, guitar (accompaniment), synth (accompaniment), and winds (accompaniment) were presented. After the ten stimuli, participants had the option to repeat the training section or to continue with the main experiment.

During the main experiment, the same procedure as in the training was used, but no feedback was given. In this section, a total of 240 stimuli were presented in random order, corresponding to 30 stimuli for each of the eight target categories.

The final section of the experiment consisted of a questionnaire regarding personal data, questions from the Gold-MSI and a debriefing that presented the achieved average detection accuracy. On average participants took 41 minutes to complete the experiments.

BEHAVIORAL ANALYSIS

Detection accuracy was determined directly from participants' responses. Following recommendations by the American Statistical Association (Wasserstein et al., 2019), we avoid assigning binary labels of "significance" to empirical results but instead provide confidence intervals of estimates where possible. Accuracies are always structured as a pair, with the first indicating the result of the target-mixture condition and the second indicating the result of the mixture-target condition. We provide mean detection accuracies followed by round brackets containing the decrease or increase through a change in presentation order.

Generalized binominal mixed-effect models (GLME; West et al., 2014) were used for statistical analyses. All mixed-effects analyses were computed in MATLAB using the *glme* function in the *Statistics and Machine Learning Toolbox* (Statistics and Machine Learning Toolbox Release 8.7, MathWorks Inc.). Our model included random intercepts for each participant and item (i.e., stimulus). All binary categorical predictors were sum-coded. To summarize the main effects and interactions, results are presented in the form of an ANOVA table, with fixed effects coefficients provided as statistical parameter (F) and probability (p), derived from the GLME models via MATLAB's *anova* function. A detailed view of the behavioral results, models and statistic evaluations are presented in the Supplementary Materials (see Tables 1–4).

FREQUENCY MICRO-MODULATION ANALYSIS

To measure the difference in FMM between the original vocal and its pitch-quantized counterparts, we evaluated the range of FMM in short time windows for unmodified vocal excerpts and pitch-quantized vocals and instruments (see Supplementary Materials Figure 3). We used a sliding window of 10 ms over the duration of the excerpt and extracted f0 via the MATLAB function *pitch*. Given that the extraction contained artifacts such as irregular fluctuations, which occurred especially in the offsets and onsets of the vocals, additional artifact suppression was applied to the extracted f0s. The artifact rejection was based on a threshold for tonal components

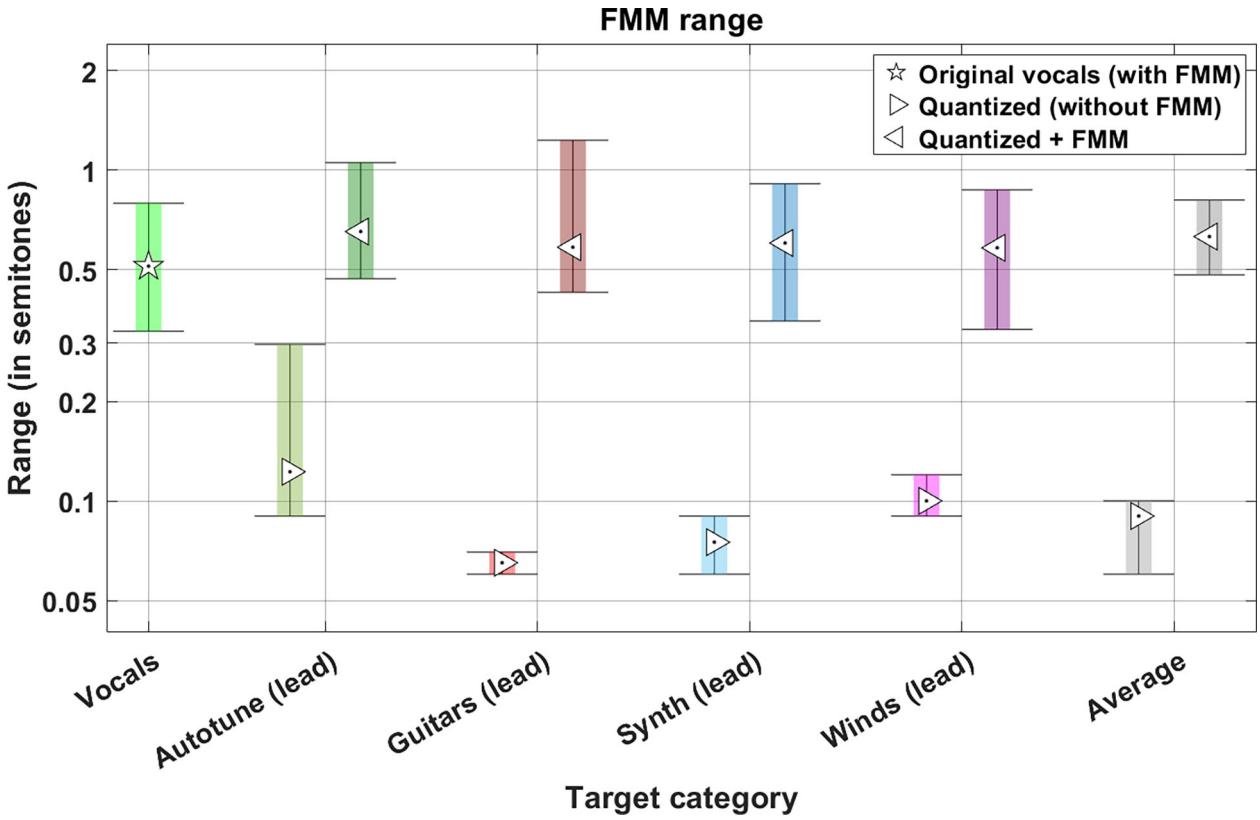

FIGURE 2. Frequency modulation analysis. To quantify the change in frequency micro-modulations between the original lead vocals, their pitch-quantized counterparts and their pitch-quantized counterparts with added frequency micro-modulation (FMM), the extracted f0 trajectories were transformed to cents and the f0 range in 100 ms time windows was evaluated. The median of the range was computed across all stimuli (30 excerpts) in each target category. "Quantized" refers to the FMM range in the autotune or melody instruments without FMM as used in Experiment 1. "Quantized + FMM" refers to the FMM range in the autotune or melody instruments with FMM as used in Experiment 2.

in the time window (harmonic ratio) as provided in the *pitch* function, excluding samples below a harmonic ratio of 75%. Additionally, a threshold for maximum f0 distance within a 100 ms sliding time window with 50% overlap was applied, excluding frequencies with a distance greater than one octave relative to the median pitch within the time window. For each excerpt and signal, the FMM range was obtained within a 100 ms sliding window by evaluating the difference in cents between the highest and lowest note. As a final step, the median across the windows was evaluated for each excerpt and signal. This excluded the relative rare time windows that contained tonal transitions. Results are presented in Figure 2 and the Supplementary Materials (see Table 5). The pitch-quantized vocal alteration showed the smallest FMM range of 0.09 semitones, whereas the FMM range of 0.51 semitones for unquantized vocals and of 0.63 semitones for the quantized alteration with FMM were considerably higher. Autotune excerpts generated directly from pitch-quantized voices showed a higher range than the excerpts generated by MIDI instruments. An additional analysis of the distance analysis between estimated f0 to perfect tempered scale tone is included in the Supplementary Materials.

Results

EXPERIMENT 1 – PITCH-QUANTIZED TARGETS

Detection accuracies of Experiment 1 are displayed in Figure 3 (for numerical values, see Supplementary Materials Table 1). A GLME included presentation order and target categories as fixed effects (see Supplementary Materials Table 2). Accuracy varied by presentation order and target category: averaged across target categories, the Target-Mixture condition yielded a higher accuracy of 88% compared to the reverse Mixture-Target condition 80% (-8%). A decline of the accuracy between the two orders was present in almost every target category but differed in size. These effects were reflected by the GLME model, with pronounced

effects for the presentation order, $F = 9.78$, $p = .002$, the targets, $F = 15.10$, $p < .001$, and the interaction between the order and targets, $F = 5.93$, $p < .001$. For readability, the following results are presented in pairs, with the first detection rate indicating the accuracy for the Target-Mixture order, and the subsequent detection rate indicating the accuracy for the Mixture-Target order. When examining the target categories, the best performing category was lead vocals with an accuracy of 99% and a minuscule decrease to 97% (-2%). The quantized voice had an accuracy of 96% but showed a decline to 87% (-9%). Targets in which the original lead vocals were replaced with an instrument showed the following accuracies: guitar from 90% to 80% (-10%), synths from 93% to 81% (-12%) and winds from 86% to 78% (-8%). Targets containing the instrumental excerpts taken from the original mixtures reached the following accuracies: guitar from 78% to 69% (-9%), the synths from 79% to 64% (-15%) and the winds from 88% to 78% (-10%).

Inspecting differences between instrument categories part of the accompaniment and those playing the main melody (guitar, synths, winds), the main melody instruments yielded clearly higher accuracies. However, the average accuracy of all main melody targets decreased considerably between presentation orders, from 89% to 79% (-10%). A similar decrease was observed for the accompanying categories with a decline from 82% to 71% (-11%). Differences between the two instrument types were analyzed using a GLME that included presentation order and instrument types as fixed effects (see Supplementary Materials Table 3). The model reflected the differences between accompaniment and main melody targets, $F = 6.95$, $p = .009$, and the influence of the presentation order, $F = 9.78$, $p < .002$. The presentation order affected each instrument in a similar way as indicated by the lack of an interaction effect between order and instrument type, $F = 0.91$, $p = .340$. The winds category behaved differently compared to the other instruments as it showed no benefit when

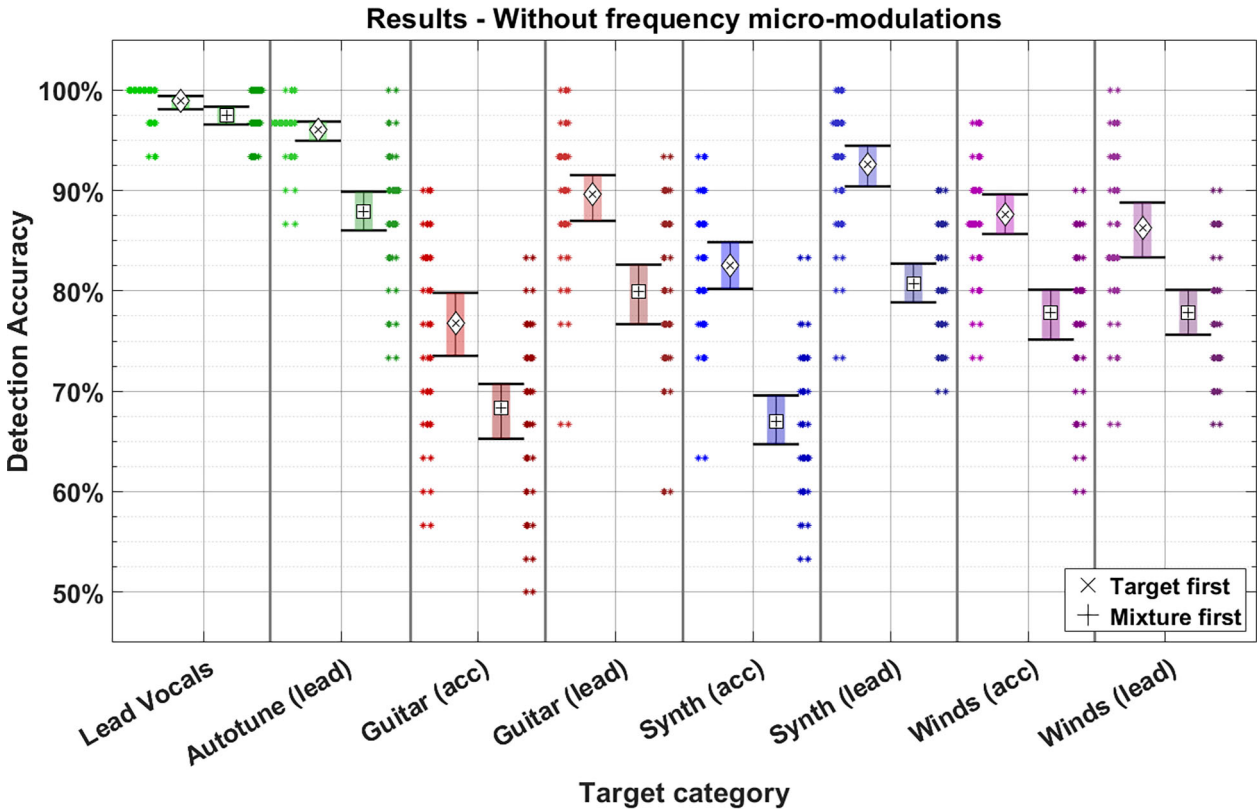

FIGURE 3. Detection accuracies for Experiment 1. Six instruments and two vocal categories were used as targets. Each instrument category was used twice either using the instrument track which was present in the excerpt (acc) or replacing the lead vocals in the excerpt by MIDI instruments using the same melody as the vocals (lead). The "x" marks the mean detection accuracy for a given target category in the presentation order "Target-Mixture." The "+" marks the mean detection accuracy for a given target category in the presentation order "Mixture-Target." Error bars indicate 95% confidence intervals. Asterisks left and right to the average of a category present average accuracies of individual participants for the given condition.

playing the main melody, but rather a minor increase when playing in the accompaniment in the Target-Mixture order (+2%) and a decrease of the same quantity in the Mixture-Target order (-2%).

In summary, there was an effect of presentation order for all targets except the original vocals. Targets were detected considerably better when the isolated target was presented first, followed by the mixture. This was also evident when target instruments that otherwise played in the accompaniment replaced the vocals in the main melody. In contrast to the original vocals with FMM, the pitch-corrected vocals without FMM showed a clear effect of presentation order. This raised the question of whether transferring FMM from vocals to instrumental signals could increase their salience. Thus, we repeated the experiment with a slight modification of the targets: we transferred the FMM of the original vocals to the respective pitch quantized vocal and main melody instrument targets.

EXPERIMENT 2 – TARGETS WITH FREQUENCY MICRO-MODULATIONS

The average detection accuracies of the second experiment are displayed in Figure 4 (for numerical values, see Supplementary Materials Table 1). A GLME included presentation order and target categories as fixed effects (see Supplementary Materials Table 2). Accuracy differed depending on the target category and order of presentation, which was also evident in our model, Order: $F = 0.41$, $p = .03$; Target: $F = 11.05$, $p < .001$; Interaction: $F = 3.49$, $p = .001$. Similar to Experiment 1, when inspecting the difference of presentation orders by averaging over target categories, the Target-Mixture condition held a higher accuracy of 90% than the Mixture-Target condition with an accuracy of 82% (-9%). When looking into the target categories, targets maintaining the original frequency trajectory of the vocals (lead vocals, autotune, and main melody instruments) revealed a clearly smaller decrease between both presentation orders than the accompanying instrument categories. This result was most pronounced in the lead vocals, which performed best with an accuracy of 98% and a decrease to 95% (-3%).

Inspecting the differences between the instrument categories playing an accompanying role and those replacing the lead vocals, the main melody instruments yielded higher accuracies. Average accuracies of all main

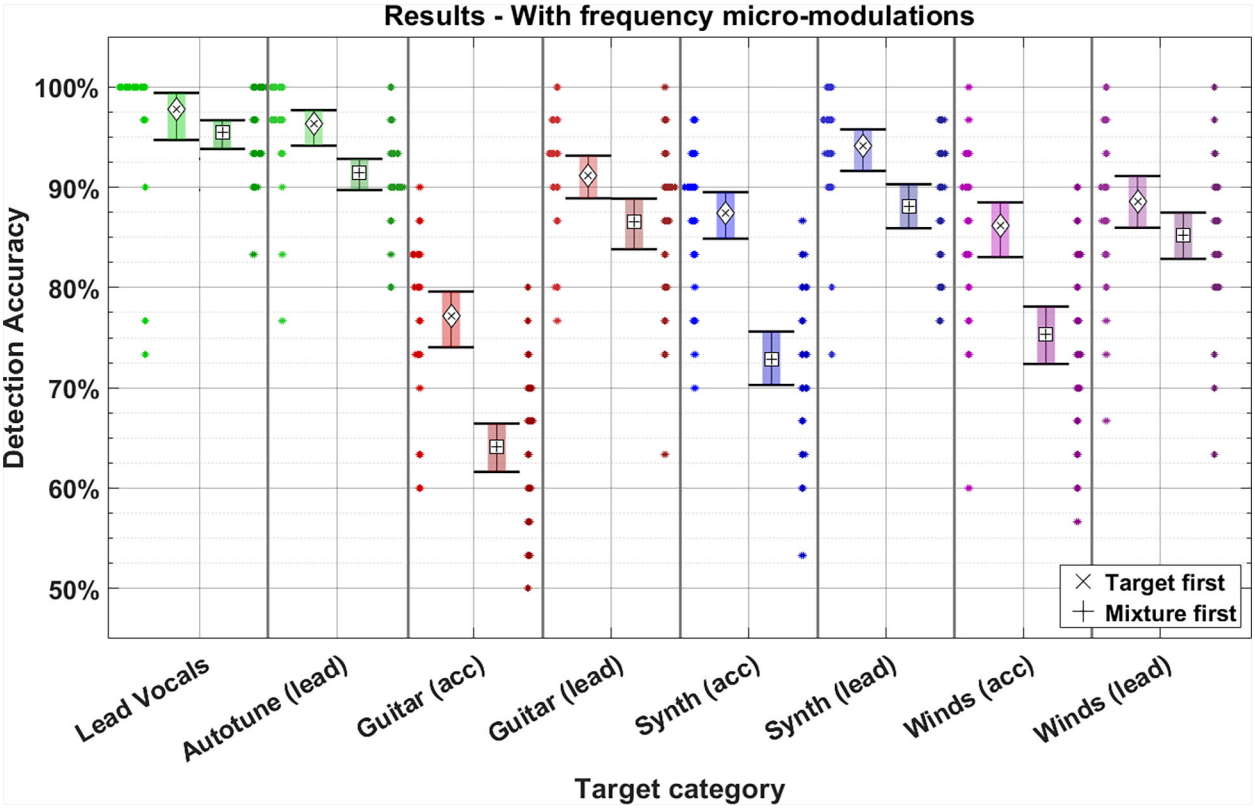

FIGURE 4. Detection accuracies for Experiment 2. Six instrumental and two vocal categories were used as targets. Each instrument category was used twice either using the instrument track which was present in the excerpt (acc) or replacing the lead vocals in the excerpt by MIDI instruments using the same melody and frequency trajectory as the vocals (lead). Graphical conventions are otherwise identical to Figure 3.

melody targets decreased across presentation orders from 91% to 87% (-4%). A larger decrease was shown for the targets part of the accompaniment with a decline from 83% to 70% (-13%). Differences between the two instrument types were analyzed using a GLME that included presentation order and musical material (accompaniment vs. main melody) as fixed effects (see Supplementary Materials Table 3). Our model reflected the differences between accompaniment and main melody targets, $F = 9.15$, $p = .003$, the influence of the presentation order, $F = 0.414$, $p = .003$, and in contrast to Experiment 1, that the presentation order affected the accompaniment and instrument targets differently by revealing an effect of the interaction between order and instrument type, $F = 47.17$, $p = .001$.

MUSICAL EXPERIENCE

Musical experience was analyzed in a questionnaire using a subset of the Gold-MSI. Nine questions regarding perceptual abilities and seven questions regarding music training were included in the questionnaire. Scores between 1 and 7 could be obtained for each question. For Experiment 1, participants reached a score of 43.4 in the perceptual abilities subscale and a score of 22.4 in the music training subscale. The correlation based on perception abilities for the Target-Mixture order was $R^2 = .001$ ($p = .89$) and for the Mixture-Target at $R^2 = .05$ ($p = .22$). Similar results were shown for the set regarding music training, with a correlation for the Target-Mixture order of $R^2 = .008$ ($p = .23$) and for the Mixture-Target order $R^2 = .054$ ($p = .20$). Regarding Experiment 2, participants reached an average score of 43.1 in the perceptual abilities' subscale and an average score of 19.4 in the music training subscale. As in Experiment 1, no notable correlations were found between the individual musical experience scores and detection accuracies. The correlation based on perception abilities for the Target-Mixture order was $R^2 = .003$ ($p = .80$) and for the Mixture-Target at $R^2 = .07$ ($p = .28$). Similar results were shown for the set regarding musical training, with a correlation for the Target-Mixture order of $R^2 = .025$ ($p = .23$) and for the Mixture-Target order $R^2 = .112$ ($p = .10$). Because we did not specifically recruit separate groups of participants with diverse degrees of musical experience, the lack of an effect of musical experience observed here was not surprising and consistent with previous research (Bürgel et al., 2021).

COMPARISON OF BOTH EXPERIMENTS

The stimuli between Experiment 1 and Experiment 2 differed only in the exclusion of FMM (Experiment 1) and the inclusion of FMM (Experiment 2) for the autotune vocals and target instruments playing the main melody. The average detection accuracy across all instruments between the two presentation orders revealed a slightly better performance in Experiment 2 with a miniscule difference of two percentage points between experiments in both presentation orders. Stimuli that remained consistent across experiments showed differences in accuracy from zero to four percentage points. Yet overall performance was similar, with an average difference between the vocals and accompanying instruments of less than one percentage point. A direct comparison of detection accuracies in both experiments for the autotune and main melody instruments is shown in Figure 5A. There were negligible differences in the Target-Mixture condition by about one percentage point. However, in the Mixture-Target condition, the autotune and melody instruments in Experiment 2 showed an enhanced detection of six percentage points compared to Experiment 1. To statistically evaluate the differences between both experiments, a GLME was utilized that included presentation order, musical role, and the different experiments as fixed effects. The model corroborated the influence of FMM (see Supplementary Materials Table 4) by indicating no interaction between presentation order and musical role when averaged across both experiments, $F = 0.62$, $p = .430$, but a three-way interaction between presentation order, musical role, and experiment, $F = 11.23$, $p < .001$. This underlines that the presence of FMM in Experiment 2 boosted performance in the otherwise difficult Mixture-Target condition of the main melody targets (see Fig. 5A). In addition, a strong correlation of $R^2 = .90$ was found between the FMM range and the order effect expressed as difference in detection accuracy of both presentation orders (see Figure 5B). Taken together, this further suggests that FMM enriches the vocals by an important factor for creating auditory salience in musical scenes.

Discussion

In the present study, we analyzed the acoustical and musical underpinnings of the lead vocals, which contribute to their role as an elevated point of auditory attention in musical mixtures (vocal salience). We investigated the influence of frequency micro-modulation (FMM) of the lead vocals and the role of the main melody in hearing out individual instruments from a mix. Specifically, participants were asked to detect cued vocals and instruments in two-second excerpts of Western popular music. To investigate the influence of attentional cues on the detection of the target, the

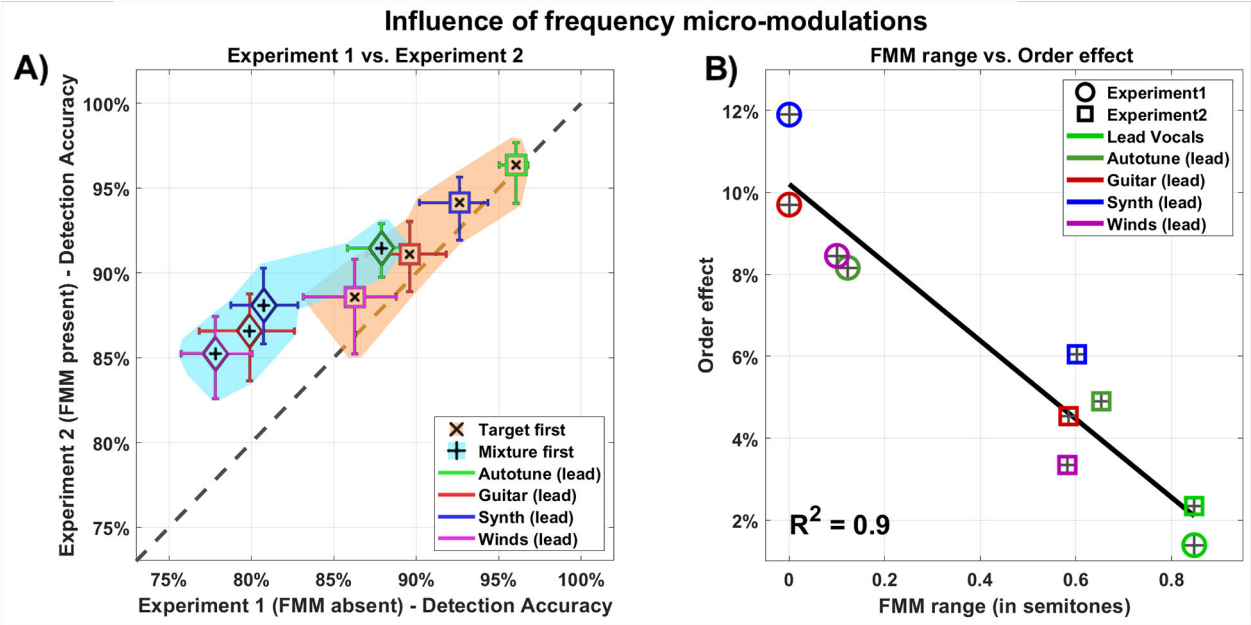

FIGURE 5. Influence of frequency micro-modulations. (A) Detection accuracy in selected conditions from Experiment 1 (*x*-axis) and Experiment 2 (*y*-axis): Two-dimensional error bars indicate 95% confidence intervals. Note that for the presented target categories, average accuracies in the "Mixture-first" conditions were significantly higher in Experiment 2 (with FMM) compared to Experiment 1 (without FMM), whereas this was not the case for "Target-first" conditions. (B) Correlation of frequency micro-modulation range and order effect: The FMM range is represented by the median range of each lead-melody target from Experiment 1 and Experiment 2. The order effect is quantified for each lead-melody target as the difference between the average detection accuracy of the "Target-first" and "Mixture-first" conditions in the respective experiment and condition.

presentation order of cue and mixture was swapped between participants, whereby the comparison between both orders revealed to which degree detection was modulated by attention (order effect). To analyze the role of the main melody for contribution to the vocal salience, instrument targets were either used in their role as part of the musical accompaniment or they were used as a replacement for the lead vocals; that is, they played the melody of the vocals. We added a vocal target category with pitch-quantized lead vocals, eliminating FMM inherent in the vocals (Experiment 1). Additionally, we repeated a modified version of the experiment in which we transferred the FMM of the lead vocals to the pitch-quantized vocals and the instruments replacing the vocals (Experiment 2).

ORDER EFFECT AND VOCAL SALIENCE

Consistent with classic studies (e.g., Bey & McAdams 2002), the presentation order of the cue played a key role in our results. When the cue preceded the mixture, listeners were able use this information to direct selective attention towards the cued signal. This resulted in higher detection rates compared to when the cue was presented subsequent to the mixture. Consistent with our previous experiments (Bürgel et al., 2021) and our hypothesis, this effect was evident in all target categories except the lead vocals, which showed only a slight decrease of accuracy when the cue was presented after the mixture. This finding highlights a unique vocal salience that enables the vocals to attract the listeners' attention, even when listening blindly into a musical scene. The present study used a different database of music excerpts compared to our previous work (Bürgel et al., 2021). The consistency of our findings across different music databases supports our general hypothesis that vocal salience in mixtures of popular music is not the result of a specific mixing strategy in music production, but rather an effect inherent in vocal signals. Previous studies have established a perceptually privileged role of the voice through the presentation of isolated voices and instruments (e.g., Agus et al., 2012; Gunji et al., 2003; Levy et al., 2001). Our present results extend this line of research by demonstrating that this effect is also present in musical mixtures.

EFFECT OF MAIN MELODY

When the guitar and synthesizer replaced the vocals as the main melody of a song, overall detection accuracy

improved, supporting our hypothesis and previous studies (Ragert et al., 2014) of a stronger perceptual salience of melody instruments over accompaniment instruments. Surprisingly, wind instruments showed no improvement whatsoever. One likely reason for this contrasting effect relates to the specific musical role of the different categories of instruments as part of the accompaniment. Whereas the guitar and synthesizer mostly played chord-based progressions in our excerpts, the winds played accompanying melodies. Consequently, the transition to the melody of the lead vocal might be a rather small change for the wind instruments, but a more drastic change of musical material for guitars and synthesizers. Nonetheless, it is important to note that the differences between the two presentation orders were still present and almost unaffected for instruments playing the vocal melodies. This implies that instruments playing the main melody are generally easier to detect, but playing the main melody does not automatically guarantee salience in a musical mixture (i.e., does not automatically attract auditory attention without a cue signal). It should be kept in mind that we used a consistent MIDI instrument for each of the individual instrument categories, which potentially may have added detection cues, although such cues would have been identical for both presentation orders. Whereas the timbre of accompanying instruments could vary between excerpts (because we used the original instruments within a song), the timbre of the three instruments playing the vocal melody did not vary. Even though we attempted to balance this aspect of experimental design by interspersing excerpts in which the vocals were replaced by instruments while the target was an accompaniment instrument, we cannot rule out that participants became accustomed to the timbre of the MIDI instruments over the duration of the experiment and implicitly memorized specific timbral properties of the MIDI instruments (Agus et al., 2010; Siedenburg & McAdams, 2018; Siedenburg & Müllensiefen, 2019).

EFFECT OF FREQUENCY MICRO-MODULATIONS

The pitch-quantized vocal category showed degradation in the Mixture-Target order, while also performing somewhat worse compared to the lead vocals in the Target-Mixture order. This suggests that excessively pitch-corrected voices do not capture listeners' attention to the same extent as more naturalistic singing voices and therefore are more likely to fuse with elements of the accompaniment in musical mixtures. This pattern of results further refutes the assumption that phonological cues are the basis of vocal salience because pitch quantization did not affect the phonological content of the vocals. One reason for the loss of attentional cues in the quantized vocals appears to be the lack of FMM, which was reduced compared to the original vocals. An acoustical analysis corroborated this interpretation by revealing a greater range of FMM for the unquantized vocals and instruments compared to their quantized counterparts that strongly correlated with the strength of the order effect. This finding is consistent with previous studies that have shown specific facilitated processing of speech with naturalistic frequency modulations, which is more intelligible compared to speech without, with decreased or exaggerated modulations (e.g., Miller et al., 2010; Wingfield et al., 1984). Furthermore, FMM has been shown to facilitate the detection of concurrently presented vowel sounds (Marin & McAdams, 1991; McAdams, 1989). Our findings extend the literature in this regard by demonstrating that the salience of vocals in musical mixtures strongly relies on frequency modulations that are present in naturalistic singing voices, helping the vocals to stand out from the mixture and attract listeners' attention.

The influence of FMM was further corroborated in our second experiment. We repeated the experiment using the same excerpts while adding the FMM of the original vocal excerpts to the instruments substituting the main melodies and quantized vocals. For the signals with artificially added FMM, our results showed a considerably reduced difference between the presentation orders in comparison to the first experiment. Interestingly, when the cue was presented before the mixture, the targets achieved very similar results across both experiments. This contradicted our hypothesis because the additional FMM did not increase overall detection but only seemed to increase the detection in the Mixture-Target order. Thus, the modulations appeared to increase the salience of the target when no prior cue was provided, drawing the attention towards the target in a similar way as seen in the lead vocals.

Curiously, even the pitch-quantized vocals with micro-modulations showed small differences between the orders of presentation, although the differences to the original vocals were supposed to be eliminated by the transfer of FMM. This result implies that although the micro-modulations make a strong contribution to vocal salience, it seems that the full salience effect may emerge from the conjunction of multiple features of the vocals. One of the features might be the pitch offset of the unaltered vocals that was eliminated by quantizing the pitch to a tempered scale. These intonation deviations occur even in professional singers (Hutchins & Campbell, 2009; Mori et al., 2004; Sundberg et al., 1996) and are an inevitable consequence of

imperfect motor controls of the voice (Hutchins et al., 2014.). Even though these deviations were unlikely to be perceived as intonation errors (Hutchins et al., 2012), it is possible that these deviations yield auditory grouping cues that let the vocals stand out of the mixture. Furthermore, singers intentionally create such deviations to add expressivity to the sound (Sundberg et al., 2013) and therefore may add an important cue to the unaltered vocals, that is lost in pitch-quantization.

More speculatively, the pitch quantization and reintroduction of pitch variation may have also altered timbral features of the vocals. Timbre is a multidimensional attribute (Siedenburg et al., 2019) that enables the discrimination and identification of sound sources (e.g., sounds from a keyboard vs. a guitar), even though they may match in other acoustic cues such as loudness and pitch. Previous studies focusing on the recognition of instruments and voices showed that the human singing voice has an advantage over other instruments supposedly based on timbre alone (Agus et al., 2012; Isnard et al., 2019; Suied et al., 2014). Voice specific cortical areas remain selective to timbre of naturalistic vocal sounds even when vocal and non-vocal sounds where matched in acoustic cues (Bélizaire et al., 2007). Further, the facilitated recognition and cortical selective was observed only for natural vocals and was absent when "chimeras," i.e., interpolations between instruments and vocals were presented (Agus et al., 2012; Agus et. al., 2017). Even though we think that in the present experiments timbre changes were subtle, if noticeable at all, this interpretation would suggest that vocal salience could be a result of the joint contribution of timbre and pitch cues in auditory scene analysis. A distortion of such joint features due to the autotuning and f0-modulation could have hindered voice-specific processing to occur, thus hindering the full salience effect to arise for our modified vocals.

In summary, in line with previous experiments, the detectability of all non-vocal instruments was affected by a change in the presentation order, whereas lead vocals were detected with similarly high accuracies in both presentation orders. This effect corroborates a unique vocal salience that automatically attracts listeners' attention. Instruments replacing vocals showed better detection accuracies compared to instruments playing as part of the musical accompaniment, but still exhibited reduced accuracy when the mixture preceded the target. Even for pitch-quantized vocals, this dependency on presentation order was evident, implying that phonological features that engage a facilitated processing of speech sounds are not sufficient to drive vocal salience. The difference between the presentation orders decreased considerably when the FMM originally present in the vocals were transferred to the instruments and pitch-quantized vocals. Overall, this also implies that excessive pitch correction may strip vocals of a unique acoustical feature that helps turning the human voice into a focal point of musical scenes.

Author Note

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

MB and KS designed the study. MB collected and analyzed the data. MB wrote a first draft of the manuscript. KS revised the manuscript. All authors contributed to the article and approved the submitted version.

This research was supported by a Freigeist Fellowship of the Volkswagen Foundation to K.S.

We thank research assistant Ghifar Aldebs for helping create the stimuli for our experiments.

Correspondence concerning this article should be addressed to Michel Bürgel, Department of Medical Physics and Acoustics, University of Oldenburg, Küpkersweg 74, 26129 Oldenburg, Germany. E-mail: michel.buergel@uol.de

References

Agus, T. R., Paquette, S., Suied, C., Pressnitzer, D., & Belin, P. (2017). Voice selectivity in the temporal voice area despite matched low-level acoustic cues. *Scientific Reports*, *7*(1), 11526. https://doi.org/10.1038/s41598-017-11684-1

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

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26. https://doi.org/10.1016/S0926-6410(01)00084-2

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312. https://doi.org/10.1038/35002078

Bélizaire, G., Fillion-Bilodeau, S., Chartrand, J.-P., Bertrand-Gauvin, C., & Belin, P (2007). Cerebral response to 'voiceness': A functional magnetic resonance imaging study. *Neuroreport*, *18*(1), 29–33. https://doi.org/10.1097/WNR.0b013e3280122718

Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception and Psychophysics*, *64*(5), 844–854. https://doi.org/10.3758/bf03194750.

Bregman, A. S., & McAdams, S. (1994). Auditory scene analysis: The perceptual organization of sound. *Journal of the Acoustical Society of America*, *95*(2), 1177–1178. https://doi.org/10.1121/1.408434

Bürgel, M., Picinali, L., & Siedenburg, K. (2021). Listening in the mix: Lead vocals robustly attract auditory attention in popular music. *Frontiers in Psychology*, *12*, 769663. https://doi.org/10.3389/fpsyg.2021.769663

Gao, Z., & Oxenham, A. J. (2022). Voice disadvantage effects in absolute and relative pitch judgments. *Journal of the Acoustical Society of America*, *151*(4), 2414–2428. https://doi.org/10.1121/10.0010123

Gunji, A., Koyama, S., Ishii, R., Levy, D., Okamoto, H., Kakigi, R., & Pantev, C. (2003). Magnetoencephalographic study of the cortical activity elicited by human voice. *Neuroscience Letters*, *348*(1), 13–16. https://doi.org/10.1016/s0304-3940(03)00640-2

Hutchins, S., & Campbell, D. (2009). Estimating the time to reach a target frequency in singing. *Annals of the New York Academy of Sciences*, *1169*, 116–120. https://doi.org/10.1111/j.1749-6632.2009.04856.x

Hutchins, S., Larrouy-Maestri, P., & Peretz, I. (2014). Singing ability is rooted in vocal-motor control of pitch. *Attention, Perception and Psychophysics*, *76*(8), 2522–2530. https://doi.org/10.3758/s13414-014-0732-1

Hutchins, S., Roquet, C., & Peretz, I. (2012). The vocal generosity effect: How bad can your singing be? *Music Perception*, *30*(2), 147–159. https://doi.org/10.1525/mp.2012.30.2.147

Isnard, V., Chastres, V., Viaud-Delmon, I., & Suied, C. (2019). The time course of auditory recognition measured with rapid sequences of short natural sounds. *Scientific Reports*, *9*(1), 8005. https://doi.org/10.1038/s41598-019-43126-5

Larrouy-Maestri, P., & Pfordresher, P. Q. (2018). Pitch perception in music: Do scoops matter? *Journal of Experimental Psychology. Human Perception and Performance*, *44*(10), 1523–1541. https://doi.org/10.1037/xhp0000550

Levy, D. A., Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli: Electrophysiological evidence. *Neuroreport*, *12*(12), 2653–2657. https://doi.org/10.1097/00001756-200108280-00013

Marin, C. M., & McAdams, S. (1991). Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width. *Journal of the Acoustical Society of America*, *89*(1), 341–351. https://doi.org/10.1121/1.400469

McAdams, S. (1989). Segregation of concurrent sounds. I: Effects of frequency modulation coherence. *Journal of the Acoustical Society of America*, *86*(6), 2148–2159. https://doi.org/10.1121/1.398475

Miller, S. E., Schlauch, R. S., & Watson, P. J. (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. *Journal of the Acoustical Society of America*, *128*(1), 435–443. https://doi.org/10.1121/1.3397384

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, *53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Mori, H., Odagiri W., Hideki., Honda, K. (2004). Transitional characteristics of fundamental frequency in singing. *Internal Congress on Acoustics* (ICA), 499–500.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS One*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642

Norman-Haignere, S. V., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A., McDermott, J. H., et al. (2022). A neural population selective for song in human auditory cortex. *Current Biology*, *32*(7), 1470-1484.e12. https://doi.org/10.1016/j.cub.2022.01.069

Ragert, M., Fairhurst, M. T., & Keller, P. E. (2014). Segregation and integration of auditory streams when listening to multi-part music. *PLOS One*, *9*(1), e84085. https://doi.org/10.1371/journal.pone.0084085

Saitou, T., Unoki, M., & Akagi, M. (2005). Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis. *Speech Communication*, *46*(3–4), 405–417. https://doi.org/10.1016/j.specom.2005.01.010

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*(3), 114–123. https://doi.org/10.1016/j.tins.2010.11.002

Signoret, C., Gaudrain, E., Tillmann, B., Grimault, N., & Perrin, F. (2011). Facilitated auditory detection for speech sounds. *Frontiers in Psychology*, *2*, 176. https://doi.org/10.3389/fpsyg.2011.00176

Siedenburg, K., & McAdams, S. (2018). Short-term recognition of timbre sequences. *Music Perception*, *36*(1), 24–39. https://doi.org/10.1525/mp.2018.36.1.24

Siedenburg, K., & Müllensiefen, D. (2019). Memory for timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Springer handbook of auditory research. Timbre: Acoustics, perception, and cognition* (Vol. *69*, pp. 87–118). Springer International Publishing. https://doi.org/10.1007/978-3-030-14832-4_4

Siedenburg, K., Saitis, C., & McAdams, S. (2019). The present, past, and future of timbre research. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Springer handbook of auditory research. Timbre: Acoustics, perception, and cognition* (Vol. *69*, pp. 1–19). Springer International Publishing. https://doi.org/10.1007/978-3-030-14832-4_1

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. https://doi.org/10.1121/1.4863659

Sundberg, J., Lã, F. M. B., & Himonides, E. (2013). Intonation and expressivity: A single case study of classical Western singing. *Journal of Voice: Official Journal of the Voice Foundation*, *27*(3), 391.e1-8. https://doi.org/10.1016/j.jvoice.2012.11.009

Sundberg, J., Prame, E., & Iwarsson, J. (1996). Replicability and accuracy of pitch patterns in professional singers. In P. J. Davis, & N. H. Fletcher (Eds.), *Vocal fold physiology: Controlling complexity and chaos* (pp. 291–306). Singular Publishing Group, Inc.

Sussman, E. S. (2017). Auditory scene analysis: An attention perspective. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2989–3000. https://doi.org/10.1044/2017_JSLHR-H-17-0041

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "$p < .05$." *The American Statistician*, *73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings: Enhanced memory for vocal melodies. *Psychological Science*, *23*(10), 1074–1078. https://doi.org/10.1177/0956797612442552

West, B. T., Welch, K. B., Galecki, A. T. (2014). *Linear mixed models*. CRC Press.

Wingfield, A., Lombardi, L., & Sokol, S. (1984). Prosodic features and the intelligibility of accelerated speech: Syntactic versus periodic segmentation. *Journal of Speech and Hearing Research*, *27*(1), 128–134. https://doi.org/10.1044/jshr.2701.128