Musical Keys Diverge in Perception but Not in Performance

Ahyeon Choi[1,*] **Woojae Cho**[2,*] **Kyogu Lee**[2]
[1]Department of Communication Sciences and Disorders, University of Iowa
[2]Department of Intelligence and Information, Seoul National University
ahyeon-choi@uiowa.edu {woojae, kglee}@snu.ac.kr
*These authors contributed equally.

Abstract

Musical key is associated with distinctive sensory qualities, yet it remains unclear whether such impressions reflect pitch height alone or notation-related key representations beyond the major–minor distinction. This study tested whether sharp and flat keys evoke different sensory impressions and whether these differences are reflected in performance. In Experiment 1, musicians rated brightness, warmth, and shape while listening to Mozart excerpts transposed across all keys and into enharmonic key pairs. Sharp-labeled keys were rated as brighter, colder, and more pointed than flat-labeled keys, even when pitch content was identical. In Experiment 2, violinists performed excerpts in the same key conditions, and acoustic features were analyzed. No reliable sharp–flat differences were observed in the acoustic measures examined here. These findings suggest that sharp–flat distinctions can shape explicit perceptual judgments beyond pitch height without necessarily producing corresponding differences in measured performance acoustics.

Keywords: Musical keys; Sensory perception; Tonal cognition; Music performance; Cross-modal correspondence

Introduction

In everyday life, music gives rise to a wide range of sensory experiences, from impressions of brightness and warmth to more abstract qualities such as shape (Kwak & Kim, 2018; Marks, 1974; Wang & Spence, 2017). These sensory dimensions contribute fundamentally to musical meaning, shaping how listeners perceive emotional qualities, atmosphere, and expressive character in music (Spence, 2020). For musicians, such sensory impressions extend beyond passive listening and are closely intertwined with active performance, motor planning, and emotional engagement (Giraldo et al., 2019; Maimon et al., 2020; Shibuya & Hasegawa, 2016). The origins of these experiences are complex, emerging from interactions among multiple musical elements, including melody, rhythm, harmony, tonality, timbre, and tempo (Bonetti & Costa, 2019; Collier & Hubbard, 2004; Fotidzis et al., 2018; Giannos et al., 2021; Timmers, 2022; Wallmark, 2019). More broadly, such experiences raise a fundamental cognitive question: to what extent are sensory impressions determined by continuous physical stimulus dimensions, and to what extent are they shaped by abstract, category-level representations.

Among these elements, musical key has long been regarded as a central factor shaping the sensory and expressive character of music. Composers such as Johann Sebastian Bach and Frédéric Chopin explored the expressive potential of all 24 major and minor keys through systematic collections encompassing all tonalities, and many composers and theorists have attributed distinct expressive or sensory qualities to individual keys. For example, C minor has frequently been described as tragic or dramatic, and Beethoven himself reportedly characterized keys using concrete expressive terms (Schindler & MacArdle, 1996; Vernon, 1942). Empirical research on key-related perception, however, has primarily focused on the distinction between major and minor modes, with major keys typically associated with positive emotions and minor keys with more negative affective qualities (Parncutt, 2014; Virtala & Tervaniemi, 2017). While this binary distinction captures robust affective differences, it does not fully account for the diversity of sensory experiences attributed to musical keys, suggesting that additional tonal factors beyond mode may play a role in shaping key-related sensory experience.

One such factor concerns the distribution of accidentals within the circle of fifths. Sharp keys are generated by ascending perfect fifths, whereas flat keys arise from descending perfect fifths (Figure 1A), a contrast that has been linked to differences in perceived tension and relaxation (Dowling, 2010). Accordingly, sharp and flat keys may evoke systematically different sensory impressions, suggesting that tonal categories defined by symbolic structure and key signature may shape sensory experience beyond their acoustic realization. From this perspective, Ruland (Ruland, 2014) proposed a perceptual continuum in which sharp keys become progressively brighter and flat keys progressively darker along the circle of fifths, and Steblin (Steblin, 1981) documented historical attempts to group keys by shared expressive qualities. However, empirical support for such key-specific associations remains mixed. For example, Powell and Dibben (Powell & Dibben, 2005) reported no reliable emotional differences across transpositions of identical musical excerpts, though their study was limited in key coverage and participant characteristics.

To address the limitations of previous work, the present study adopted an experimental framework designed to disentangle key-related sensory representations from pitch-based effects. While transposing identical musical material across keys allows all tonal conditions to be evaluated, it also introduces a fundamental confound: pitch height varies systematically with key, and pitch height is strongly associated with sensory impressions such as brightness, warmth, and shape. These crossmodal correspondences between acoustic and sensory dimensions are well established, with higher pitches typ-

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

ically perceived as brighter, colder, and more pointed, and lower pitches as darker, warmer, and rounder (Eitan & Rothschild, 2011; Kussner, 2014; Kwak & Kim, 2018; Kwak et al., 2020; Marks, 1974; Wang & Spence, 2017).

Because pitch–sensation mappings are approximately linear within the audible range, isolating key-specific effects requires explicit control of pitch. The present study therefore used the stimulus design shown in Figure 1A to contrast pitch-varying and pitch-controlled conditions across two experiments. In the pitch-varying condition, identical excerpts were transposed across all keys, allowing pitch height to covary with key. In the pitch-controlled condition, stimuli were restricted to enharmonic keys. In modern equal temperament, enharmonic keys such as C♯ major and D♭ major have identical pitch content but different key signatures, making them useful for separating pitch-based effects from notation-related key associations.

Experiment 1 examined perceptual judgments of musical keys using this two-session structure (Figure 1B). In Session 1, identical excerpts were transposed into all keys, providing a baseline for pitch-driven sensory trends. In Session 2, stimuli were restricted to enharmonic key pairs, enabling direct comparisons between sharp- and flat-labeled keys while strictly controlling pitch content. Together, these sessions tested whether sharp–flat distinctions in sensory perception persist beyond continuous pitch-based mappings.

Building on the perceptual findings of Experiment 1, Experiment 2 extended the same conceptual framework to musical performance (Figure 1B). Professional violinists performed stimuli based on the same pitch-varying and pitch-controlled structure, and their performances were recorded for acoustic analysis. Together, the two experiments test whether sharp–flat differences reflect pitch-based sensory mappings alone, or whether notation-related key representations shape perceptual judgments without producing corresponding differences in the acoustic measures examined here.

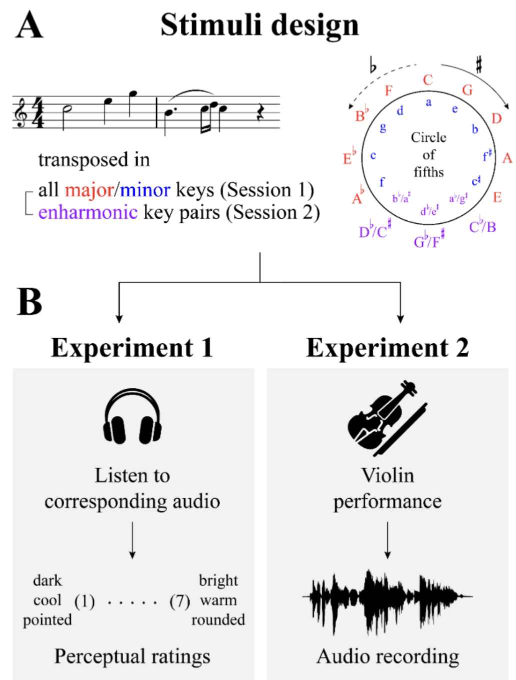

Figure 1: Overview of the stimuli design and experimental procedure. (A) Musical excerpts were transposed into all major and minor keys in Session 1 and into enharmonic key pairs in Session 2. (B) Experiment 1 assessed perceptual ratings while participants listened to the corresponding audio, whereas Experiment 2 recorded violin performances for acoustic analysis.

Experiment 1

Method

Participants A total of 30 music-major students participated in the experiment. The mean age of the participants was 22.2 years (SD = 2.2), and the mean duration of formal musical training was 8.8 years (SD = 4.3). Among the participants, 14 majored in piano, 11 in composition or music theory, and 5 in string instruments (violin, viola, or cello). Prior to participation, all procedures were explained in detail, and written informed consent was obtained from each participant. The study was reviewed and approved by the Institutional Review Board of Seoul National University (IRB No. 2007/002-022).

Stimuli Experiment 1 consisted of two sessions conducted with the same group of participants, following the stimuli design summarized in Figure 1A. All stimuli were excerpts from piano sonatas by Wolfgang Amadeus Mozart that clearly established a tonality within the first eight measures. Mozart's works were selected because they provide a clear sense of tonality while maintaining an appropriate length and musical complexity for the present task. To minimize potential effects of familiarity with specific key signatures, only sonatas originally composed in C major and A minor—keys without accidentals—were used. All stimuli were digitally transposed, and musical scores and corresponding audio were presented simultaneously.

The two sessions differed only in the selection of musical excerpts and their transposition strategy. In Session 1, the first movements of Mozart's Piano Sonata K. 545 (C major) and K. 310 (A minor) were transposed into all 24 keys (12 major and 12 minor). Including enharmonic equivalents (e.g., B/C♭, C♯/D♭, F♯/G♭ for major; g♯/a♭, a♯/b♭, d♯/e♭ for minor), this yielded a total of 30 stimuli (15 major and 15 minor). Keys were ordered by ascending tonic pitch to maintain a comparable pitch range across stimuli.

In Session 2, six additional Mozart sonata movements were used: the first and third movements of three C-major piano sonatas (K. 270, K. 309, and K. 330). These excerpts were transposed exclusively into six enharmonic key pairs (B/C♭, C♯/D♭, F♯/G♭), yielding 36 stimuli. Across both sessions, stimulus presentation was randomized, with the constraint that enharmonic pairs sharing identical pitch content were not presented consecutively. This design enabled sharp–flat comparisons under strict pitch control, isolating notational effects from continuous pitch variation.

Procedure Before starting the main experiment, participants completed a demographic and background questionnaire reporting their gender, age, hearing and vision status (normal or corrected-to-normal), duration of formal musical training, and primary instrument. In both Session 1 and Session 2, participants listened to each musical excerpt while viewing the corresponding score and were asked to evaluate the sensory impressions evoked by the key of the music. For every trial, participants rated three sensory dimensions—Brightness, Warmth, and Shape—using a 7-point Likert scale (1 = dark, cold, or pointed; 7 = bright, warm, or round). The same rating procedure was applied consistently across both sessions.

Participants were allowed a short break of approximately 30–60 seconds between Session 1 and Session 2. The total duration of the entire experiment was approximately 40 minutes. After completing both sessions, participants took part in a brief post-experiment interview. They were invited to provide open-ended responses to questions such as: "Do you associate particular feelings or qualities with specific keys?", "Did you notice any differences in how sharp or flat keys felt during the experiment?", and "What sensory images come to mind when you think of sharps or flats?" All responses were recorded and summarized for qualitative reference.

Data analysis For Session 1, mean ratings for each sensory dimension were computed for every key separately for major and minor modes. To examine the relationship between tonic pitch height and sensory perception, linear regressions were fitted across all keys for each dimension (Brightness, Warmth, and Shape), and coefficients of determination (r^2) were used to quantify the strength of pitch–sensation correspondences. Regression trends were examined to assess whether sharp and flat keys exhibited systematic deviations beyond pitch height.

For Session 2, ratings were averaged across trials for each participant and notation type (sharp vs. flat). Paired t-tests were conducted to directly compare sensory ratings between sharp and flat notations for each dimension. To control for multiple comparisons, Bonferroni correction was applied, and statistical significance was evaluated at an adjusted alpha level. For both sessions, warmth and shape scales were reversed for visualization and interpretability in terms of cross-modal congruence such that higher values indicate colder and more pointed sensations.

Results

Session 1 Session 1 examined whether sensory impressions of musical keys varied systematically with tonic pitch height and whether sharp- and flat-labeled keys exhibited distinct perceptual trends beyond pitch height. For each sensory dimension—brightness, warmth, and shape—a linear regression was fitted across all keys, and coefficients of determination (r^2) were used to quantify the strength of the pitch–sensation relationship. Figure 2 summarizes the results for all three dimensions.

Across dimensions, sensory ratings showed strong linear associations with tonic pitch height. Higher-pitched keys were generally rated as brighter, colder, and more pointed. A broad mode-related difference was also observed: major keys tended to be perceived as brighter, warmer, and rounder, whereas minor keys were rated as darker, colder, and more pointed. This major–minor contrast was expected based on previous findings and was not the primary focus of the present analysis.

Crucially, within both major and minor modes, sharp- and flat-labeled keys showed consistent separations across the pitch continuum. Brightness ratings (Figure 2A) exhibited robust linearity for both major (r^2 = .847) and minor keys (r^2 = .890), with sharp keys generally rated as brighter than flat keys, including across enharmonic pairs with identical pitch content. Warmth ratings (Figure 2B) also demonstrated

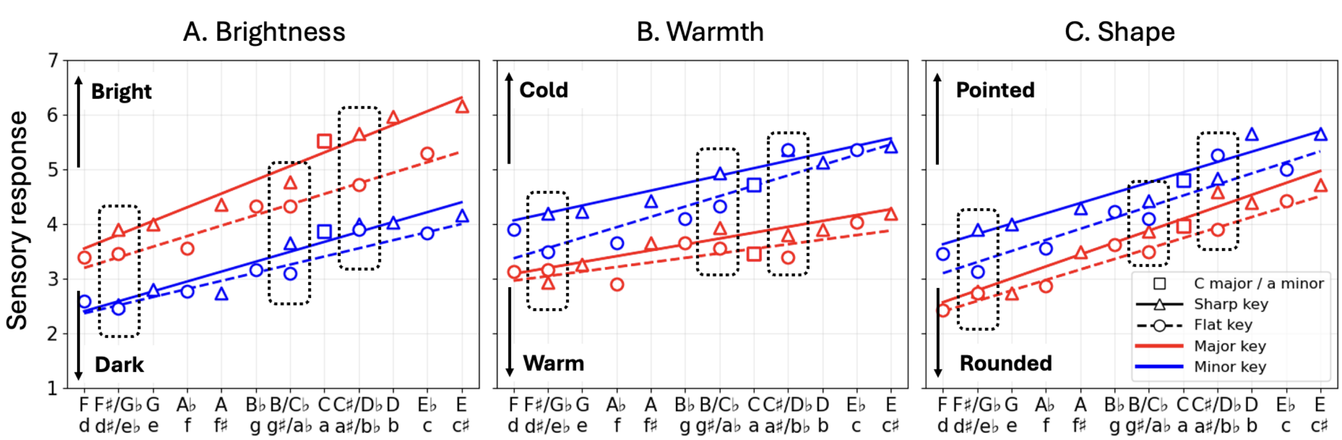

Figure 2: Mean sensory ratings across all keys for (A) Brightness, (B) Warmth, and (C) Shape, shown for major and minor keys. Brightness uses the original 7-point scale (1 = dark, 7 = bright). For visualization, Warmth and Shape scales were reversed such that higher values indicate colder and more pointed sensations (original scales: 1 = cold / pointed, 7 = warm / round).

substantial linearity with pitch height (r^2 = .711 for major; r^2 = .796 for minor), with sharp keys tending to be rated as colder than flat keys. Shape ratings (Figure 2C), indexed such that higher values indicate greater pointedness, showed the strongest linear trends among the three dimensions (r^2 = .938 for major; r^2 = .846 for minor), with sharp keys generally rated as more pointed than flat keys.

Taken together, Session 1 indicates that sensory impressions of musical keys were strongly structured by continuous pitch-based mappings, while also showing sharp–flat separations that could not be reduced entirely to pitch height. These patterns motivated the pitch-controlled comparison using enharmonic key pairs in Session 2.

Session 2 The paired t-tests revealed significant differences between sharp- and flat-key ratings for all three sensory dimensions (Table 1). Despite identical pitch content, participants rated sharp-labeled keys as brighter, colder, and more pointed than flat-labeled keys. These results indicate that sharp–flat notation systematically modulated sensory impressions of identical musical material.

Table 1: Comparison of mean sensory ratings between sharp and flat keys (paired t-test, N = 30). Higher values indicate brighter, colder, and more pointed responses.

	Sharp M	Flat M	t	Sig.
Brightness	4.68	4.44	3.74	***
Warmth (reversed)	3.75	3.43	2.96	**
Shape (reversed)	4.16	3.62	5.45	***

$^{**}p < .01$, $^{***}p < .001$. Bonferroni correction applied.

Post-experiment Interviews In post-experiment interviews, many participants reported sensory associations consistent with the rating results. Sharp keys were most often described using terms such as "bright," "piercing," or "cutting," whereas flat keys were more frequently associated with descriptors such as "warm," "heavy," or "pastoral." Some participants also noted that these impressions became more pronounced as the number of sharps or flats increased. String-instrument majors sometimes attributed these impressions to differences in timbre across keys or to variations in tactile feedback during performance. These observations were used as qualitative context rather than as independent evidence.

Experiment 2

Experiment 2 tested whether the perceptual sharp–flat differences observed in Experiment 1 were reflected in violin performance. Violinists performed stimuli based on the same pitch-varying and pitch-controlled design, and the resulting recordings were analyzed acoustically.

Method

Participants A new group of participants, independent from those in Experiment 1, took part in Experiment 2. All participants were trained musicians with a primary specialization in violin performance. Initially, 19 violinists were recruited and assigned to one of two sessions (10 in Session 1, 9 in Session 2) to reduce the time and physical demands of performance recording. However, data from two participants from Session 1 were excluded due to audio recording issues, resulting in a final sample of 17 participants. The mean age of the participants was 23.88 years (SD = 6.00), and the mean duration of formal musical training was 10.94 years (SD = 4.77). Prior to participation, all procedures were explained in detail, and written informed consent was obtained from each participant. The study was reviewed and approved by the Institutional Review Board of Seoul National University (IRB No. 2304/001-01).

Stimuli The overall stimulus structure and key-transposition strategy were identical to those used in Experiment 1, with the primary difference being the use of violin repertoire instead of piano music. All stimuli consisted of excerpts from violin sonatas by Wolfgang Amadeus Mozart that clearly established a tonality within the first eight measures and were suitable for short performance-based trials.

Experiment 2 consisted of two sessions. In Session 1, one major-key excerpt (Mozart Violin Sonata K. 373 in C major) and one minor-key excerpt (Mozart Violin Sonata K. 402 in A minor) were each transposed into 15 keys, yielding a total of 30 stimuli (15 major, 15 minor). Key ordering and pitch-range constraints followed the same principles as in Experiment 1, ensuring comparable and playable pitch ranges across transpositions.

In Session 2, six excerpts originally composed in C major were drawn from three violin sonatas (K. 296, K. 303, and K. 403; movements 1 and 3) and transposed into six enharmonic key pairs, yielding 36 stimuli. Final notes were minimally edited when necessary to produce perceptually convincing cadences. In both sessions, stimulus order was fully randomized, enabling direct comparison of sharp and flat notations under strictly matched pitch content.

Procedure The experimental procedure closely followed that of Experiment 1, with modifications appropriate for active musical performance. Before starting the main experiment, participants completed the same demographic and background questionnaire used in Experiment 1. Participants then tuned their instruments to a reference pitch of A = 440 Hz.

In both Session 1 and Session 2, musical scores were presented visually on a screen, and participants were instructed to perform each excerpt in synchrony with a metronome preroll to 80 beats per minute. Each excerpt required approximately 20 seconds to perform at the prescribed tempo. All performances were recorded using a condenser microphone (Neumann, U87) and an audio interface (Focusrite, Scarlett 8i6), with recordings captured in Logic Pro X. Individual trials were subsequently segmented and edited for acoustic analysis.

Data analysis To examine whether key-related perceptual differences were reflected in musical performance, acoustic features were extracted from the recorded violin performances. Prior to feature extraction, silent segments were removed from each audio file to ensure that analyses reflected only active sound production. Four acoustic features were then computed over time for each trial: fundamental frequency (F0), spectral centroid, spectral slope, and spectral flux.

These features were selected as acoustic descriptors related to perceptual dimensions such as brightness, warmth, and shape that were examined in Experiment 1, without assuming a one-to-one mapping between acoustic measures and subjective experience. Median F0 was used as a robust index of pitch height, providing a general acoustic reference aligned with the pitch-based structure observed in Session 1 of Experiment 1. Spectral centroid and spectral slope were included to characterize overall spectral energy distribution associated with timbral brightness and warmth, while spectral flux captured time-varying spectral changes related to articulatory dynamics in performance.

For each trial, a single representative value was computed for each feature across time. Median values were used for F0 to reduce sensitivity to outliers, whereas mean values were computed for spectral centroid, spectral slope, and spectral flux. These trial-level summary measures were then used for statistical analysis and visualization. Feature values were plotted as a function of key, following the same visualization conventions used in Experiment 1, to allow direct comparison between perceptual ratings and performance-based acoustic patterns.

Results

Session 1 Session 1 examined whether the perceptual sharp–flat patterns observed in Experiment 1 were reflected in the acoustic properties of violin performance when pitch height varied systematically across keys (Figure 3). Because the major and minor excerpts differed in their original pitch ranges, comparisons between modes were not interpreted as key effects. Instead, analyses focused on whether sharp- and flat-labeled keys exhibited systematic deviations beyond pitch height.

As shown in Figure 3A, median F0 closely tracked tonic pitch height across all keys, exhibiting near-perfect linearity (r^2 = .988–.994). This pattern was consistent across sharp- and flat-labeled keys in both modes, with no meaningful separation between them. These results confirm that pitch height was realized with high precision across keys, providing a baseline against which potential notation-related deviations could be evaluated.

Spectral centroid likewise increased monotonically with key pitch height (Figure 3B), showing strong linear associations across all key groups (r^2 = .934–.980). Spectral slope demonstrated similarly robust linear trends with pitch height (Figure 3C; r^2 = .971–.989). In both features, sharp- and flat-labeled keys followed overlapping trajectories, indicating that spectral characteristics related to brightness and warmth were primarily governed by pitch-based structure rather than by sharp–flat notation.

In contrast, spectral flux exhibited weaker and more variable relationships with key pitch height (Figure 3D; r^2 = .684–.829). No consistent differences were observed between sharp- and flat-labeled keys, suggesting that dynamic articulatory features were less tightly constrained by tonal structure and more strongly influenced by performance variability.

Taken together, Session 1 indicates that violin performance acoustics were largely governed by pitch height and closely tracked the pitch-based structure of the key. Unlike the robust sharp–flat distinctions observed in perceptual ratings in Experiment 1, no corresponding differentiation emerged in the acoustic features examined here when pitch height covaried with key. This pattern motivated the pitch-controlled comparison using enharmonic key pairs in Session 2.

Session 2 Session 2 tested whether sharp- and flat-labeled keys yielded distinct acoustic realizations in violin performance when pitch content was strictly controlled using enharmonic key pairs. For each acoustic feature, paired com-

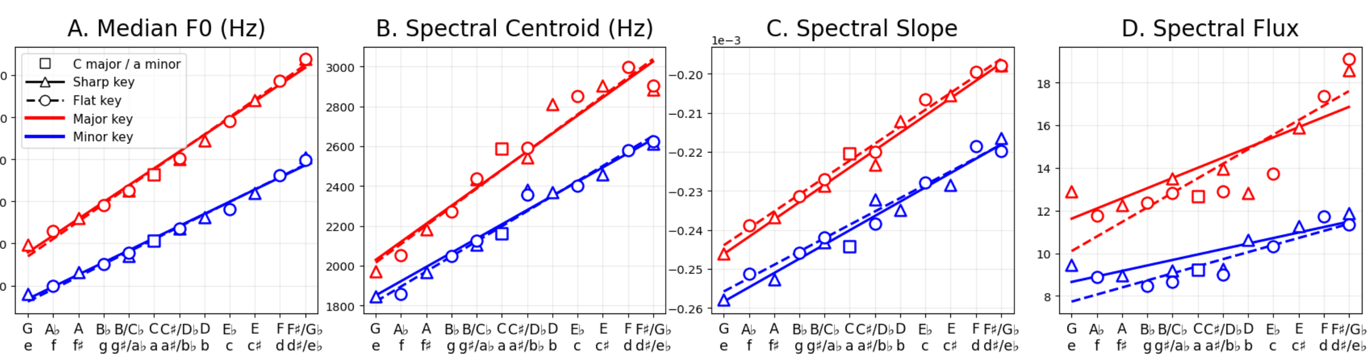

Figure 3: Mean acoustic feature values across all keys in Experiment 2 Session 1, shown separately for major and minor excerpts. Panels show (A) median fundamental frequency (F0), (B) mean spectral centroid, (C) mean spectral slope, and (D) mean spectral flux.

parisons were conducted using Wilcoxon signed-rank tests (Table 2).

Table 2: Paired comparisons of acoustic features between sharp and flat keys in Experiment 2 Session 2 (Wilcoxon signed-rank tests).

Feature	Sharp *M*	Flat *M*	*W*	*p*
F0	348.22	350.24	734	.089
Centroid	2430.14	2536.63	1608	.077
Slope	-2.15×10^{-4}	-2.11×10^{-4}	1670	.129
Flux	8.70	8.29	1640	.101

No acoustic feature showed a statistically significant difference between sharp- and flat-labeled keys. Median F0 did not differ reliably between notations, indicating that pitch height was closely matched across enharmonic pairs. Spectral centroid likewise showed no significant sharp–flat difference, suggesting that performance-based correlates of brightness were not systematically modulated by notation. Similarly, neither spectral slope nor spectral flux exhibited significant differences between sharp- and flat-labeled keys, indicating that spectral distribution and dynamic articulatory characteristics were comparable across notations. Together with Session 1, these findings indicate that the acoustic features examined here did not systematically differ between sharp and flat keys when pitch content was controlled.

General Discussion

The present study examined whether sensory qualities attributed to musical keys arise solely from continuous pitch-based mappings, or whether notation-related key representations contribute to perceptual judgments beyond pitch height. Across two experiments, we observed different patterns in perception and performance: sharp–flat differences were reliable in explicit sensory ratings, but they were not reflected in the acoustic features of violin performance examined here.

In Experiment 1, sharp- and flat-labeled keys evoked systematically different sensory impressions. Sharp keys were rated as brighter, colder, and more pointed, whereas flat keys were rated as darker, warmer, and rounder. These differences appeared both as trends across pitch height and, more critically, in enharmonic key pairs with identical pitch content. This pattern suggests that sensory impressions of musical key are not fully determined by acoustic pitch height. Instead, they may also reflect notation-related associations linked to learned tonal categories, key signatures, and symbolic musical structure. At the same time, the strong linear relationships between pitch height and sensory ratings confirm that continuous pitch-based mappings remain a major factor in shaping key-related sensory experience.

Experiment 2 tested whether these perceptual sharp–flat differences were mirrored in acoustic features of violin performance. Unlike the perceptual ratings, median F0, spectral centroid, spectral slope, and spectral flux showed no reliable sharp–flat differences, even when pitch content was controlled using enharmonic key pairs. These results suggest that the notation-related associations observed in explicit perceptual judgments did not produce corresponding differences in the summary acoustic measures analyzed here. Rather than indicating that such associations are absent from performance altogether, this finding suggests that they may be reduced, constrained, or expressed in ways not captured by the present acoustic features.

Several alternative interpretations should be considered. First, because participants in Experiment 1 listened while viewing the score, the perceptual effects may partly reflect score-based listening rather than auditory processing alone. The visual forms of the sharp sign (#) and flat sign (♭) may also have contributed to the results: the angular sharp sign and the curved flat sign could support sound–shape associations that influence explicit ratings. Second, musicians, especially those trained on non-fretted instruments, may rely on intonation schemas that do not always correspond strictly to equal temperament. Thus, the observed perceptual effects may reflect learned notation-related and performance-related associations rather than purely abstract key categories.

The present study also has limits on generality. The perceptual and performance tasks were completed by different groups of expert musicians, so the design does not test whether individuals who report stronger sharp–flat sensory associations also produce different acoustic realizations. In addition, the findings cannot determine whether similar effects would be observed in non-musicians or listeners with less notation training. Finally, Experiment 2 focused on a limited set of summary acoustic features. Other aspects of performance, such as local intonation, articulation, dynamics, timing, or vibrato, may reveal more subtle notation-related differences and should be examined in future work.

Taken together, the findings suggest that sharp–flat distinctions can shape explicit sensory judgments of musical key beyond pitch height, but that these differences are not necessarily expressed in the acoustic features measured during violin performance. Musical key therefore provides a useful domain for examining how learned symbolic systems can influence perceptual experience while remaining only partially reflected, or not reflected, in measurable performance output.

Conclusion

This study examined whether sensory qualities attributed to musical keys are driven solely by pitch height or also shaped by notation-related key representations. Perceptual judgments revealed systematic sharp–flat differences even under strict pitch control, whereas the acoustic features examined in violin performance showed no corresponding differences. These findings suggest that sharp–flat distinctions can influence explicit sensory judgments of musical key beyond pitch-based mappings, but are not necessarily reflected in the acoustic measures analyzed here. More broadly, musical key provides a useful case for examining how learned symbolic systems shape perceptual experience in expert music cognition.

Acknowledgments

This work was partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) [No. RS-2024-00461617, 50%; No. RS-2025-24683892, 45%] and by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [No. RS-2021-II211343, 5%].

References

Bonetti, L., & Costa, M. (2019). Musical mode and visual-spatial cross-modal associations in infants and adults. *Musicae Scientiae*, *23*(1), 50–68.

Collier, W. G., & Hubbard, T. L. (2004). Musical scales and brightness evaluations: Effects of pitch, direction, and scale mode. *Musicae Scientiae*, *8*(2), 151–173.

Dowling, W. J. (2010). Qualia as intervening variables in the understanding of music cognition. *Musica Humana*, *2*(1), 1–20.

Eitan, Z., & Rothschild, I. (2011). How music touches: Musical parameters and listeners' audio-tactile metaphorical mappings. *Psychology of Music*, *39*(4), 449–467.

Fotidzis, T. S., Moon, H., Steele, J. R., & Magne, C. L. (2018). Cross-modal priming effect of rhythm on visual word recognition and its relationships to music aptitude and reading achievement. *Brain sciences*, *8*(12), 210.

Giannos, K., Athanasopoulos, G., & Cambouropoulos, E. (2021). Cross-modal associations between harmonic dissonance and visual roughness. *Music & Science*, *4*, 20592043211055484.

Giraldo, S., Waddell, G., Nou, I., Ortega, A., Mayor, O., Perez, A., Williamon, A., & Ramirez, R. (2019). Automatic assessment of tone quality in violin music performance. *Frontiers in Psychology*, *10*, 334.

Kussner, M. (2014). *Shape, drawing and gesture: Cross-modal mappings of sound and music* [Doctoral dissertation, King's College London].

Kwak, Y., & Kim, C.-Y. (2018). Cross-modal correspondence between acoustic feature and shape [(in Korean)]. *Korean Journal of Cognitive and Biological Psychology*, *30*(3), 269–275.

Kwak, Y., Nam, H., & Kim, C.-Y. (2020). Cross-modal correspondence between speech sound and visual shape influencing perceptual representation of shape: The role of articulation and pitch. *Multisensory Research*, *33*(6), 589–592.

Maimon, N. B., Lamy, D., & Eitan, Z. (2020). Crossmodal correspondence between tonal hierarchy and visual brightness: Associating syntactic structure and perceptual dimensions across modalities. *Multisensory Research*, *33*(8), 805–836.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American Journal of Psychology*, *87*(1–2), 173–188.

Parncutt, R. (2014). The emotional connotations of major versus minor tonality: One or more origins? *Musicae Scientiae*, *18*(3), 336–342.

Powell, J., & Dibben, N. (2005). Key–mood association: A self-perpetuating myth. *Musicae Scientiae*, *9*(2), 296–309.

Ruland, H. (2014). *Expanding tonal awareness: A musical exploration of the evolution of consciousness from ancient tone systems to new tonalities guided by the monochord*. Rudolf Steiner Press.

Schindler, A., & MacArdle, D. W. (1996). *Beethoven as I knew him*. Courier Corporation.

Shibuya, K., & Hasegawa, N. (2016). How violinists alter sound volume and tempo when expressing bright and dark timbres. *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 138–143.

Spence, C. (2020). Assessing the role of emotional mediation in explaining crossmodal correspondences involving musical stimuli. *Multisensory Research*, *33*(1), 1–29.

Steblin, R. K. (1981). *Key characteristics in the 18th and early 19th centuries: A historical approach*. University of Illinois Press.

Timmers, R. (2022). Cross-modality and embodiment of tempo and timing. *The Oxford handbook of time in music*, 215–234.

Vernon, P. E. (1942). The individuality of keys. *The Musical Times*, *83*(1190), 105–107.

Virtala, P., & Tervaniemi, M. (2017). Neurocognition of major–minor and consonance–dissonance. *Music Perception: An Interdisciplinary Journal*, *34*(4), 394–397.

Wallmark, Z. (2019). Semantic crosstalk in timbre perception. *Music & Science*, *2*, 2059204319846617.

Wang, Q. J., & Spence, C. (2017). The role of pitch and tempo in sound–temperature crossmodal correspondences. *Multisensory Research*, *30*(3–5), 315–316.