DOI: 10.61782/fa.2025.0329

EFFECTS OF DIFFERENT AUDITORY MONITORING CONDITIONS ON VOCAL INTONATION ACCURACY

Agnieszka Paula Pietrzak[1*]
[1] Institute of Radioelectronics and Multimedia Technology, Warsaw University of Technology, Poland

ABSTRACT

Precise intonation is crucial for ensemble singing performances, particularly when singers record their parts separately under varying monitoring conditions. This study investigated the intonation accuracy of five female choristers from an academic choir under three common feedback setups: speaker, open headphones, and closed headphones. Recordings were made in an anechoic chamber, where participants sang a perfect fifth (while the reference sound was still audible) and a major third (after the reference sound had ended). Intonation accuracy was measured in cents as the difference between the sung pitch and the expected frequency. Results showed that, for the perfect fifth (harmonic interval), using closed headphones resulted in the smallest median error (11 cents), suggesting that greater isolation may enhance pitch focus. However, for the major third (melodic interval), open headphones produced the lowest error (16 cents), indicating that too much isolation can lead to pitch overestimation when reference tones are no longer audible. Across both tasks, speaker monitoring exhibited the highest overall error values. Analysis of variance (ANOVA) confirmed statistically significant differences between these monitoring types. These findings suggest that selecting appropriate auditory feedback in both recording and rehearsal settings can optimize singers' intonation.

Keywords: *vocal intonation, auditory monitoring, headphones, pitch accuracy*

[*]***Corresponding author***: *agnieszka.pietrzak@pw.edu.pl*

1. INTRODUCTION

The quality of ensemble or choir singing depends largely on the precise vocal intonation of its participants. Even small pitch discrepancies, initially below the threshold of casual perception, can accumulate and become audible tuning problem in vocal performances [1, 2].
While such issues can typically be resolved during in-person rehearsals, technological developments have expanded the contexts in which choral or ensemble recordings now take place, including situations where individual singers record their parts separately. From modern "virtual choir" projects to remote recording sessions, singers increasingly rely on various forms of auditory monitoring to maintain pitch accuracy with previously recorded tracks or external reference tones. Considering the advancements in modern recording techniques, understanding how auditory feedback mechanisms can influence pitch accuracy becomes important.
Research on vocal pitch control and intonation has long acknowledged the importance of auditory feedback [3]. Singers rely on both external cues (e.g., accompaniment, other vocal parts, or pre-recorded tracks) and internal cues (e.g., bone conduction, proprioceptive and tactile sensations) to guide their pitch production. When these cues are altered, the singer's ability to accurately judge and produce pitch can worsen [4,5]. Different types of headphones or speaker setups used as auditory monitoring can be a source of this kind of alteration. Previous studies have shown that closed-back headphones, for instance, can help isolate a vocal part and reduce external noise, thereby promoting more focused listening; however, they may also amplify the singer's own internal sounds (e.g., breathing, bone-conducted resonance), potentially leading to overestimation of one's pitch when an external reference is removed [6]. By contrast, open-back headphones allow for a more natural blend of external and internal sound, but they may also introduce ambient noise or bleed, which can

interfere with the clear perception of reference pitches. Speaker monitoring, while common in studio settings, can cause subtle pitch discrepancies depending on room acoustics or speaker placement, and may not offer the same level of isolation as headphones.

To objectively assess the impact of different monitoring methods on intonation, precise measurement metrics are necessary. Pitch accuracy in vocal performance is typically quantified in cents, with one semitone equivalent to 100 cents. Previous studies have shown that well-trained singers generally maintain intonation errors within the range of 10-20 cents [1,3]. In many cases, professional vocalists can achieve even lower error margins, often in the vicinity of 10–15 cents, reflecting their advanced auditory-motor control and extensive training. In contrast, less experienced singers or those operating in less controlled environments may exhibit errors exceeding 30 cents [6].

Based on these observations, this study aims to compare the effects of closed-back headphones, open-back headphones, and speaker monitoring on vocal intonation across both continuous (harmonic) and recalled (melodic) pitch tasks.

2. METHOD

This study investigated the influence of three monitoring conditions: open-back headphones (Beyerdynamic DT 990 Pro), closed-back headphones (Audio-Technica ATH-M50X), and speaker monitoring (JBL 4208), on singers' vocal intonation accuracy. The recordings were conducted with six female volunteers (aged 20–35) who were active members of the Warsaw University of Technology Academic Choir, but not professional vocalists. Of these six participants, one produced only the reference sounds, while the remaining five performed the target singing tasks. The small and homogeneous sample is a limitation of this study, which may restrict the generalizability of the findings, but only female voices were chosen, to limit the number of factors for analysis.

To generate reference material, one chorister recorded designated pitches that were initially presented by pure-tone sinusoids at known frequencies (e.g., 294 Hz, 392 Hz, 440 Hz). From these recordings, samples that demonstrated the most stable pitch and minimal errors were then selected for use as the reference stimuli. This approach was chosen over a fully synthesized tone to better reflect authentic vocal conditions.

All participants stood 15–20 cm from a Neumann U87 microphone, centrally placed in the anechoic chamber. The output level for each monitoring setup (open-back headphones, closed-back headphones, and speaker) was calibrated to 70 dB SPL at the participant's position. While the anechoic chamber was used to ensure experimental control, it may not appropriately represent the typical performance settings encountered in everyday practice.

The experiment comprised two tasks. In the first, singers performed a perfect fifth above a sung reference, thereby creating a harmonic interval. The perfect fifth was chosen for its high degree of consonance and harmonic stability. Three reference frequencies were used for this task (294 Hz, 349 Hz, and 392 Hz) and the singers were asked to sing the perfect fifths in just intonation (3:2 ratio). For each frequency there were five repetitions under each monitoring condition. In the second task, participants heard two sung notes forming a perfect fifth and were asked to sing a major third (5:4 ratio) after the reference tones ended, thus producing the interval melodically and in isolation. In contrast to a perfect fifth, which is a very stable consonant, the major third is generally more sensitive to intonation discrepancies, which could make the second task harder to perform. The reference fifths were sung from 330 Hz, 392 Hz, and 440 Hz and likewise repeated five times for each monitoring condition.

All resulting audio files were tracked in Cakewalk, exported as .wav files, and then analyzed in Python with the Librosa library. Fundamental frequency (F0) was extracted using the probabilistic YIN (pYIN) algorithm [7]. Given the typical pitch variability at note onsets and releases, only the 2-second portion deemed most stable, starting 0.5 seconds after the note attack, was used for subsequent measurements.

Intonation accuracy was assessed by calculating pitch error in cents, as the difference between the singer's average pitch and the expected frequency for the specified interval. Cents were used to compare pitch error at different absolute frequencies. Finally, a one-way analysis of variance (ANOVA) was performed to assess whether the selected monitoring condition exerted a statistically significant effect on intonation accuracy across both the harmonic (perfect fifth) and melodic (major third) intervals.

3. RESULTS

For both considered tasks: singing a perfect fifth while a reference tone was still audible (harmonic interval) and singing a complimentary major third after two reference tones had ceased (melodic interval) – performed analysis revealed statistically significant differences in performance across the three monitoring conditions.

For the harmonic interval of perfect fifth (Fig. 1), results showed the smallest median pitch error for closed headphones (11 cents). Open-back headphones yielded a slightly higher but still relatively small median pitch error (13 cents), while speaker monitoring produced the largest pitch deviations (18 cents). Additionally, the distribution of individual values was most consistent (i.e., narrowest) under closed headphones, while speaker monitoring had the widest variability among participants.

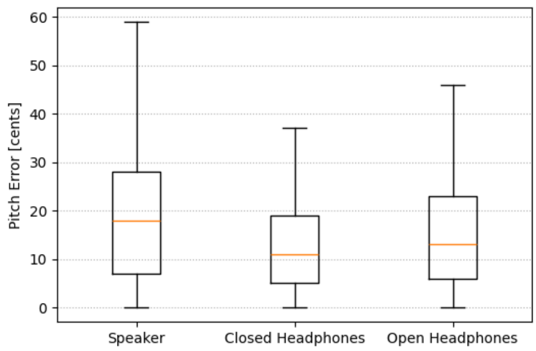

Figure 1. Box plot illustrating the pitch error (in cents) for a harmonic fifth sung under three different monitoring conditions (speaker, closed-back headphones, open-back headphones).

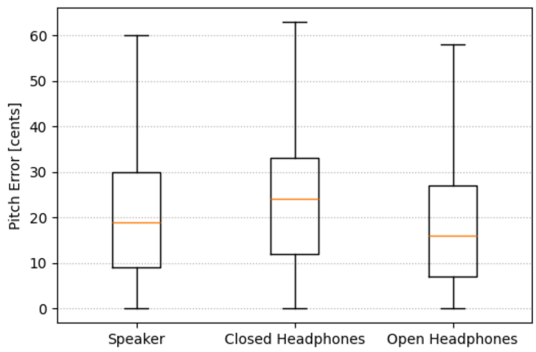

Figure 2. Box plot illustrating the pitch error (in cents) for a melodic third sung under three different monitoring conditions (speaker, closed-back headphones, open-back headphones).

When singing the major third without a continuously sounding reference, the patterns shifted (Fig. 2). Open-back headphones gave the lowest pitch error (16 cents), followed by the speaker setup (19 cents), and then closed headphones (24 cents). In this melodic context, the increased isolation provided by closed headphones appears to have contributed to overestimation of pitch, likely because participants relied primarily on their own vocal feedback once the reference was no longer audible.

A combined analysis of recorded intervals showed that the highest overall median error (19 cents) was observed for the speaker monitoring, followed by closed-back headphones (18 cents) and open-back headphones (16 cents). One-way analyses of variance (ANOVAs) for each interval type (perfect fifth: $F = 1243.4$, $p < 0.05$, major third: $F = 250.0$, $p < 0.05$), as well as for the combined data ($F = 343.4$, $p < 0.05$), confirmed that the differences between monitoring conditions were statistically significant.

Obtained results suggest that while closed-back headphones minimized pitch discrepancies for the harmonic (perfect fifth) interval, open-back headphones were more beneficial when singers had to recall pitch internally for the melodic (major third) task. Speaker monitoring consistently showed higher deviations in pitch accuracy compared to either headphone approach.

In addition to the statistically significant differences observed, the results may indicate, that even though the median differences between conditions are small, ranging from roughly 2 to 7 cents, they can be meaningful in professional contexts. Research indicates that trained musicians and critical listening environments can detect pitch differences as small as 2–5 cents [8]. Therefore, even these modest variations in pitch error may have practical implications for vocal performance quality, particularly in high-fidelity recording and live settings.

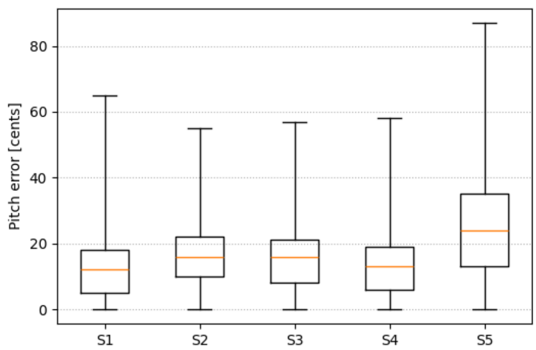

Figure 3. Box plot showing the combined pitch error distributions for singers S1–S5 across both intervals, highlighting individual variations in intonation accuracy.

Some insight can be also obtained by examination of individual singers performance. Figure 3 illustrates the singer-specific aggregated pitch error distributions, combined for both the harmonic (perfect fifth) and melodic (major third) tasks. Overall, S1, S2, S3, and S4 show moderate deviations, typically under 20 cents, indicating stable intonation. S5, however, displays a higher median error (24 cents) with the maximum exceeding 80 cents, suggesting greater difficulty controlling pitch. S5's broader error may suggest a tendency toward larger pitch mismatches, possibly due to individual factors such as technique, reliance on external cues, or comfort with recalling intervals.

4. CONCLUSIONS

This study demonstrates that the choice of monitoring method can affect choristers' ability to achieve precise intonation, especially when considering whether reference pitches are continuously present (harmonic tasks) or must be recalled (melodic tasks). Closed-back headphones produced the lowest intonation errors for a harmonic perfect fifth, suggesting that stronger isolation helps singers maintain focus on a reference pitch. However, for the melodic major third, sung without an ongoing reference, open-back headphones yielded the best results, implying that overly isolated conditions may lead to pitch overestimation once external cues are removed. Speaker monitoring, meanwhile, was associated with generally higher error values, showing its limitations for precise pitch accuracy during individual recordings.

These findings are in line with existing literature, which reports that well-trained singers typically maintain intonation errors in the range of 10–20 cents [1,4], with more experienced vocalists sometimes achieving errors as low as 10–15 cents [5,6]. In contrast, less-experienced singers may exhibit errors exceeding 30 cents. Trained musicians and professionals can detect pitch differences as small as 2–5 cents [8], therefore the pitch error differences observed in our study confirm that monitoring conditions can significantly influence pitch accuracy.

While this study offers some insight to the effects of monitoring conditions on intonation accuracy, it also has some limitations. The small, homogeneous sample, consisting only of non-professional, female choristers, and the controlled environment of an anechoic chamber may restrict the generalizability of these results to more natural performance settings. Moreover, the focus on only two intervals (perfect fifth and major third) simplifies the complex reality of vocal and choral singing. Future research should consider larger, more diverse participant groups and explore a broader range of intervals and musical contexts. Investigating alternative monitoring conditions (such as semi-open headphones or in-ear monitors) and testing in more acoustically realistic environments could give further information on how auditory feedback affects intonation.

5. REFERENCES

[1] S. Ternström and J. Sundberg, "Intonation precision of choir singers," *J. Acoust. Soc. Am.*, vol. 84, no. 1, pp. 59–69, 1988.

[2] H. Cuesta, E. Gómez Gutiérrez, A. Martorell Domínguez, and F. Loáiciga, "Analysis of intonation in unison choir singing," 2018.

[3] J. Sundberg, *The Science of the Singing Voice*, 1st ed. DeKalb, IL: Northern Illinois University Press, 1987.

[4] T. D. Rossing, J. Sundberg, and S. Ternström, "Acoustic comparison of voice use in solo and choir singing," *J. Acoust. Soc. Am.*, vol. 79, no. 6, pp. 1975-1981 (1986).

[5] D. Mürbe, F. Pabst, G. Hofmann, and J. Sundberg, "Significance of auditory and kinesthetic feedback to singers' pitch control," *Journal of Voice*, vol. 16, no. 1, pp. 44-51 (2002).

[6] D. Mürbe, F. Pabst, G. Hofmann, and J. Sundberg, "Effects of a professional solo singer education on auditory and kinesthetic feedback—a longitudinal study of singers' pitch control," *Journal of Voice*, vol. 18, no. 2, pp. 236–241, 2004.

[7] M. Mauch and S. M. Dixon, "pYIN: A Fundamental Frequency Estimator Using Probabilistic Inference," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* (ICASSP), 2014, pp. 313–317.

[8] C. Micheyl, K. Delhommeau, X. Perrot, and A. J. Oxenham, "Influence of musical and psychoacoustical training on pitch discrimination," *Hearing Research*, vol. 219, no. 1–2, pp. 36–47, 2006.