2025 IMEKO TC26 International Conference on
Metrology for Archaeology and Cultural Heritage
Bergamo, Italy, October 15-17, 2025

Medieval church bell sound generation for vibroacoustic landscape studies

Marco Casazza[1], Fabrizio Barone[1]

[1.] *University of Salerno, Department of Medicine, Surgery and Dentistry "Scuola Medica Salernitana", Baronissi, Italy (mcasazza@unisa.it; fbarone@unisa.it)*

***Abstract* – This paper presents a method for generating synthetic bell sounds from in situ recordings of historical church bells, aimed at heritage-oriented metrological analysis and immersive applications. Starting from a real recording of a 13th century bell, still operating in Salerno Cathedral (Campania Region, Italy), the signal is denoised using spectral subtraction with adaptive control. Bell strikes are detected through RMS envelope analysis, and the tonal structure is extracted via FFT-based peak detection. A synthetic signal is generated by preserving both the amplitude proportions and phase information of the original spectral peaks. The resulting waveform replicates the temporal, spectral, and perceptual features of the original bell while eliminating environmental noise. A spectral comparison confirms the high coherence between the denoised and artificial signals. The proposed synthetic source is suitable for spatial audio testing, acoustic metrology, and the reconstruction of historical soundscapes, offering repeatability and fidelity for applications in heritage science, including Wave Field Synthesis and immersive cultural experiences.**

I. INTRODUCTION

Artificial sounds can be used as reference sources with known characteristics for the measurement, analysis, and interpretation of acoustic phenomena [1]. Their controlled and repeatable nature makes them particularly valuable in metrological applications, enabling the objective evaluation of perceptual and spatial sound attributes under reproducible conditions. The application of artificial sound sources can be especially useful within the fields of musical acoustics and cultural heritage studies in relation to psychoacoustics and perceptual studies, wave field synthesis (WFS) for the reconstruction of musical instrument sounds, archaeo-acoustics and soundscape studies, as well as a support to immersive storytelling in archaeology.

In psychoacoustic investigations, artificial sound sources are employed to study human auditory perception, including pitch discrimination, spatial localization, and scene analysis. Through synthesized stimuli, researchers can isolate and manipulate acoustic features, thereby allowing for controlled perceptual experiments and calibration of auditory models. Such methods contribute to the refinement of sound quality metrics and the validation of perceptual indicators for complex listening environments [2].

Wave Field Synthesis, a spatial audio rendering technique based on Huygens' principle, is based on the use of loudspeakers or loudspeaker arrays to recreate sound fields, that are spatially coherent across extended areas. This technology has proven essential in the evaluation of spatial fidelity in acoustic reconstructions and in the development of immersive audio systems for performance spaces and virtual environments [3]. From a metrological perspective, WFS enables repeatable sound field generation for in situ or laboratory testing of acoustic responses in heritage spaces.

The field of archaeo-acoustics benefits from artificial sound sources through their use in the aural reconstruction of historic environments. By combining architectural models with physically-based sound propagation simulations and controlled emission, researchers can evaluate the acoustic properties of cultural heritage structures, even when partially destroyed or inaccessible. Studies, such as the analysis of the Maior Ecclesia of Cluny, exemplify how artificial stimuli and calibrated measurement techniques can support hypotheses on liturgical soundscapes and ritual practices [4].

Finally, immersive storytelling in cultural heritage contexts increasingly relies on artificial sound sources embedded in virtual and augmented reality systems. These technologies are designed not only for educational or curatorial purposes but also as tools for perceptual validation of reconstructions. Recent systematic reviews highlight the centrality of sound in shaping users' spatial presence and emotional engagement within digitally mediated heritage experiences [5]. In these contexts, artificial sound sources serve both as a design element and as an object of metrological investigation to assess the realism and impact of the audio component.

This contribution aims to describe the logical and mathematical procedure behind a script, used to analyze the sound of a historical church bell, measured in the environment and, after its denoising, to generate the sound of a historical church bell as a characterized source with specific spectral peaks or features. The purpose of such an artificial source, constituting a novel application in the context of metrology applied to cultural heritage, is the potential use of such a characterized source as a part of a measure chain to characterize experimentally the historical vibroacoustic landscape of bell sounds located in historical bell towers. This is especially useful in the case of bell towers, where bells were removed and cannot be found, providing a potential way to increase the standardization in immersive vibroacoustic heritage studies and applications.

II. MATERIALS AND METHODS

A. Tonal resynthesis

A tonal resynthesis procedure was developed to reconstruct the acoustic signal of a medieval church bell from a noisy recording. The method is based on five steps: noise reduction, transient detection, spectral analysis, tonal synthesis with real amplitude envelope, and rhythmic reconstruction.

Let x(t) be the original noisy signal and n(t) a short recording of background noise. Both signals are sampled with the same frequency. Both time-domain signals are transformed into the time-frequency domain via the short-time Fourier transform (STFT) [6]:

$$X(t,f) = STFT\{x(t)\} \quad (1)$$

$$N(t,f) = STFT\{n(t)\} \quad (2)$$

Then, a spectral subtraction technique is applied [7]:

$$|X_{clean}(t,f)| = max(|X(t,f)| - \alpha \cdot \bar{N}(f), 0) \quad (3)$$

where $\bar{N}(f)$ is the average noise spectrum and α is a tuning parameter, used to guarantee a sufficient amplification of the resulting cleaned signal. The clean signal is reconstructed as the inverse short time Fourier transform:

$$x_{clean}(t) = ISTFT\left(|X_{clean}(t,f)| \cdot e^{j\angle X(t,f)}\right) \quad (4)$$

After performing the background noise reduction, the transient detection is performed with the purpose of determining the bell strike time. The root-mean-square (RMS) envelope of the denoised signal is computed over a sliding window of length Δ (e.g.: 50 ms):

$$e_{rms}(t) = \sqrt{\frac{1}{\Delta} \sum_{k=t-\frac{\Delta}{2}}^{t+\frac{\Delta}{2}} x(k)} \quad (5)$$

Then, bell strike times, t_k, as peaks of the envelope, when it crosses an adaptive threshold, given also a minimum temporal interval between two strikes, as commonly done in transient-based audio analysis [8]:

$$e_{rms}(t_k) = \mu_e + 2\sigma_e \quad (6)$$

$$(t_{k+1} - t_k) > 0.5 \text{ s} \quad (7)$$

In parallel, the reference sound (i.e., the medieval church bell sound) spectral features are considered. In particular, the clean signal frequency spectral content is estimated via discrete Fourier transform (DFT). Then, a certain number of frequency peaks, N, is extracted (e.g., 15 peaks) based on their prominence, eventually considering avoiding the frequencies above a certain threshold, f_{thresh}. These peaks are ranked in ascending order, to ensure the fundamental frequency is prioritized:

$$\{f_i\}_{i=1}^{N} = \text{sort}\big((peaks|X(f)|) \cap f_{thresh}\big) \quad (8)$$

The tonal re-synthesis of the bell tolls, based on the analyzed parameters, is reconstructed using a sum of sinusoids, modulated by the real envelope $e_k(t)$ extracted from the original signal:

$$s_k(\text{t}) = \sum_{i=1}^{N} A_i \cdot e_k(t) \cdot \sin(2\pi f_i(t) + \phi_i) \quad (9)$$

where:

$$A_i = \frac{p_i}{\max(p)} \quad (10)$$

$$\phi_i \in [0; 2\pi] \quad (11)$$

with A_i as normalized spectral peak magnitude, ϕ_i as random phase and $e_k(t)$ being interpolated from the original RMS envelope segment.

To simulate the impulsive excitation caused by the clapper, a Gaussian noise burst is added at the onset:

$$s_k(t) \leftarrow s_k(t) + 0.4 \cdot w(t) \cdot \eta(t) \quad (12)$$

$$\phi_i \in [0; 2\pi] \quad (13)$$

with $w(t) = Gaussian$ and $\eta(t) = white\ noise$.

The synthesized signal is constructed by temporal alignment of a certain number, K, of reconstructed strikes:

$$s_{final}(\text{t}) = \sum_{\text{k}=1}^{K} s_k(t - t_k) \quad (14)$$

The signal is globally normalized to avoid clipping:

$$s_{final}(t) \leftarrow \frac{s_{final}(\text{t})}{max|s_{final}(\text{t})|} \quad (15)$$

Finally, the signal is exported as a high-resolution .wav file, with an output frequency of 44,100 Hz (equivalent to that of the recorded church bell sound).

B. Church bell sound sampling and analysis

The acoustic signal of a historical church bell was recorded in an open urban environment using a calibrated measurement setup with known sampling frequency (44.1 kHz). To isolate the vibrational tonal content of the bell from the ambient background noise, a separate recording of the environmental noise was acquired under identical conditions, without the bell striking. The bell sound used for such a purpose is the one of the oldest bells of Salerno cathedral, casted in the 13th century by an anonymous bell maker, producing a fundamental tone of B^3.

The signals, acquired as uncompressed *.wav files, were, then, elaborated through a tailored algorithm developed in MATLAB™ R2024b to produce a final audio file containing the sound of an artificial bell with spectral features analogous to the real (denoised) bell.

The two time-domain signals — denoted as x(t) for the bell and n(t) for the background — were converted into the time-frequency domain using the Short-Time Fourier Transform (STFT), following Allen and Rabiner's formulations [6]. A spectral subtraction method was applied to remove the average spectral magnitude of the noise from the signal's spectrum, with a tunable scaling parameter α to balance preservation and clarity of the tonal components. The clean signal was reconstructed using the inverse STFT, preserving the original phase spectrum.

The algorithm, then, applies a transient to the denoised signal to detect the onset of individual bell strikes. The root-mean-square (RMS) envelope was calculated over a sliding window (typically 50 ms), and strike times were identified as the peaks of this envelope exceeding an adaptive threshold. A minimum time spacing of 0.5 seconds between consecutive peaks was enforced to prevent false detections.

The spectral profile of the bell was characterized by computing the discrete Fourier transform (DFT) of the denoised signal and identifying the most prominent frequency peaks, excluding components above a defined threshold. The frequencies were sorted in ascending order and the corresponding amplitudes were normalized with respect to the strongest component.

The original phase values associated with each detected peak were also extracted from the complex spectrum. These phases were preserved for later synthesis to retain the perceived timbral quality and temporal signature of the original bell sound.

The combined information — timing of strikes, amplitudes, frequencies, and phases — was used in the tonal re-synthesis phase, ensuring a metrologically meaningful artificial signal compatible with spatial audio reconstruction and immersive heritage scenarios.

III. RESULTS AND DISCUSSION

A. Results

Fig. 1 illustrates the identification of bell strike onsets based on the root-mean-square envelope of the denoised signal. The peaks detected above the adaptive threshold allowed the temporal segmentation of individual impulses, establishing the temporal pattern for the tonal resynthesis. In this specific case, a regular inter-strike interval was confirmed, with a mean separation of approximately 2.7 seconds between subsequent bell strikes.

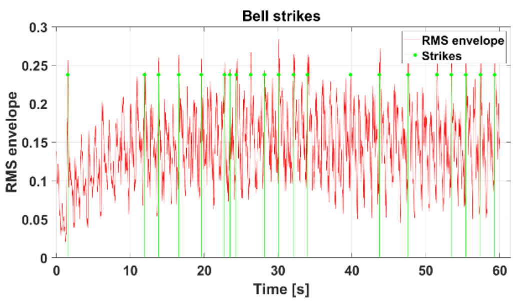

Fig. 1. Bell strikes identification for the calculation of bell strike mean time

Fig. 2 compares the original recorded signal (including environmental noise), the denoised signal obtained after spectral subtraction, and the synthetic bell signal generated using the tonal resynthesis algorithm. The synthetic waveform preserves the harmonic structure and decay characteristics of the original source while eliminating background noise and preserving a clean temporal signature. The spectral quality of the bell sound — perceived pitch, richness of partials, and impulse profile — were perceptually retained, supporting the use of the synthesized version as a proxy source for metrological and immersive audio applications.

.

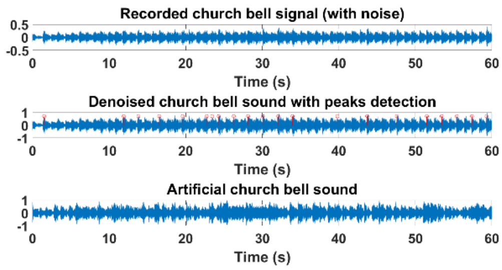

Fig. 2. Comparison between original recorded church bell signals (with environmental noise included), denoised temporal sequence and artificial church bell sound produced through the MATLAB algorithm

Fig. 3 highlights the high degree of spectral coherence between the artificial bell and the denoised church bell sound, with the main tonal peaks (including the fundamental) accurately reconstructed.

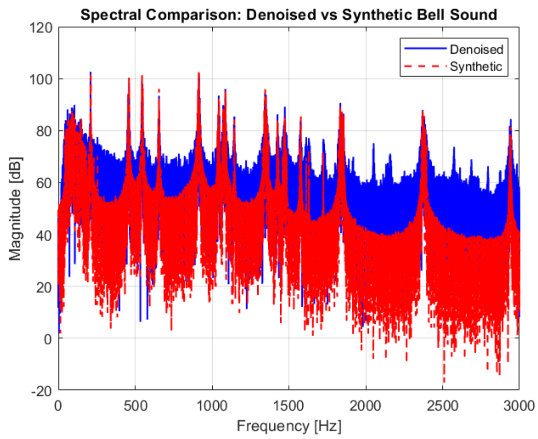

Fig. 3. Superposition of artificial bell sound and denoised church bell sound spectra, exhibiting a high degree of spectral coherence.

Additionally, the use of real phase information from the original FFT spectrum ensures that the reconstructed waveform aligns perceptually with the historical acoustic signature of the original bell.

These results demonstrate the feasibility of constructing an artificial bell source with defined and repeatable characteristics from in situ recordings, enabling applications in spatial audio reproduction, soundscape studies, and heritage-based metrological frameworks.

B. Discussion

The tonal synthesis method proposed in this work has demonstrated high compatibility with Wave Field Synthesis (WFS) applications in scenarios involving a single bell source. In WFS, the sound field generated by a virtual source is recreated through an array of loudspeakers by accurately simulating wavefronts in space. A key requirement is the availability of a controllable, clean, and temporally coherent signal associated with a well-defined spatial position [3].

The resynthesized bell signal produced by our algorithm fulfills these criteria. First, the tonal content is reconstructed from the prominent frequency components of the original signal, preserving perceptually salient harmonics. Second, the signal envelope is shaped using the real amplitude decay extracted from the input recording, which leads to a temporally realistic behavior. Third, the sequence of strikes is clearly segmented and temporally ordered, enabling effective positioning in time and space through a single point source.

Therefore, the synthesized output can be directly used as a virtual source in WFS rendering engines, provided that its spatial coordinates are defined and fed to the rendering layer. The absence of spatial directivity or phase constraints makes this implementation ideal for static and omnidirectional sources, such as isolated tower bells or test signals for spatial audio calibration.

However, several limitations arise if the method is extended to simulate multiple bell sources or distributed emission systems. The most critical limitation is the lack of phase coherence across separately synthesized sources. Since each tonal component is currently assigned a random phase, interference patterns may emerge when multiple sources are combined in the WFS rendering stage, potentially degrading the spatial perception or introducing coloration [4,5].

Additionally, the method does not account for source directivity patterns, which are relevant when simulating bells of different types or directional loudspeakers. The assumption of omnidirectionality holds only for simple cases and may need to be relaxed in more complex WFS scenarios. Finally, the system currently lacks temporal synchronization and spatial metadata management, which would be essential to orchestrate multiple synthetic signals in a unified virtual scene.

To address these issues, future work may focus on extending the method with phase-controlled synthesis, spatial tagging, and directional modeling. These enhancements would allow the system to scale towards immersive auditory environments, realistic bell towers, or historically accurate reconstructions of urban acoustic landscapes.

IV. CONCLUSIONS

This study demonstrated the feasibility and metrological relevance of generating synthetic bell sounds based on real historical church bell recordings. Through a carefully structured procedure, including noise reduction, transient detection, spectral peak analysis, and tonal resynthesis with original phase preservation, an artificial bell signal was produced that accurately reflects the vibrational and perceptual characteristics of the source bell.

The resulting waveform, tested on a 13th century bell from the Salerno Cathedral, preserved both the spectral structure and temporal behavior of the original, while eliminating environmental noise and introducing reproducibility. The comparison between the denoised and synthetic signals confirmed the high spectral coherence and the preservation of perceptually salient components.

The synthesized signal is particularly suited for applications in vibroacoustic heritage metrological research, immersive soundscapes, and spatial audio testing, including Wave Field Synthesis (WFS). Its structured construction, based on physically meaningful features, ensures traceability, standardization potential, and relevance in metrological chains focused on acoustic reconstruction of lost or inaccessible sources.

Future developments will explore the extension of the method to multi-source environments, phase synchronization across synthetic elements, and integration with spatial metadata for more complex virtual reconstructions of historical soundscapes.

REFERENCES

[1] Kawahara, H., and Yatabe, K., "Safeguarding Test Signals for Acoustic Measurement Using Arbitrary Sounds: Measuring Impulse Response by Playing Music," *Acoustical Science and Technology*, Vol. 43, No. 3, 2022, pp. 209–212. https://doi.org/10.1250/ast.43.209

[2] Schneider, A., Ed., "Studies in Musical Acoustics and Psychoacoustics," Springer International Publishing, Cham, 2017. https://doi.org/10.1007/978-3-319-47292-8

[3] Ziemer, T., "Wave Field Synthesis," *Psychoacoustic Music Sound Field Synthesis*, Vol. 7, Springer International Publishing, Cham, 2020, pp. 203–243. https://doi.org/10.1007/978-3-030-23033-3_8

[4] Suárez, R., Alonso, A., and Sendra, J. J., "Archaeoacoustics of Intangible Cultural Heritage: The Sound of the Maior Ecclesia of Cluny," *Journal of Cultural Heritage*, Vol. 19, 2016, pp. 567–572. https://doi.org/10.1016/j.culher.2015.12.003

[5] Privitera, A. G., Fontana, F., and Geronazzo, M.,

"The Role of Audio in Immersive Storytelling: A Systematic Review in Cultural Heritage," *Multimedia Tools and Applications*, 2024. https://doi.org/10.1007/s11042-024-19288-4

[6] Allen, J. B., and Rabiner, L. R., "A Unified Approach to Short-Time Fourier Analysis and Synthesis," *Proceedings of the IEEE*, Vol. 65, No. 11, 1977, pp. 1558–1564. https://doi.org/10.1109/PROC.1977.10770

[7] Boll, S., "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 27, No. 2, 1979, pp. 113–120. https://doi.org/10.1109/TASSP.1979.1163209

[8] Peeters, G., and Deruty, E., "Sound Indexing Using Morphological Description," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 3, 2010, pp. 675–687. https://doi.org/10.1109/TASL.2009.2038809