24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

AN AURALIZATION ETUDE FOR VOCAL ENSEMBLES USING ODEON

Thomas Burns

Senior Principal Engineer, Starkey Hearing Technologies, 6600 Washington Ave S., Eden Prairie, MN, USA
email: thomas.burns@starkey.com

Odeon v14 is used to auralize a large (18,000 m^3) reverberant space. The impulse response of each source is computed and convolved with an anechoic recording of an individual vocalist singing a musical excerpt alone; SATB excerpts are combined to auralize an entire ensemble. The directivity function of the vocalists is acquired from a 3D measurement of a KEMAR HATS; the same HATS is used to acquire the binaural HRIR for the receiver. Various choral arrangements and Odeon settings will be available at the poster session for binaural listening via high-fidelity, in-ear monitors.

1. Introduction

The motivation for this study was to explore a new process to generate realistic auralizations of vocal ensembles. Standard anechoic vocal recordings available within Odeon include a single, operatic soprano; alto, tenor, or bass recordings aren't available. Wenger Corporation of Owatonna, MN has recordings [1] of an eighty-member vocal ensemble acquired in a large anechoic chamber at 3M Corporation, St. Paul, MN. Convolving that 80-member recording using a single source/receiver position in Odeon has certain implications based on simplicity and computational ease. My desire was to: 1) create source/receiver positions that correspond with a choral configuration, 2) convolve their impulse responses with anechoic recordings of individual singers, 3) apply directivity functions to sources and binaural HRIR's to the receiver as measured from a KEMAR HATS, and 4) combine multiple singers into a single Odeon auralization.

2. Archimetrics of the space

The model used in this study was based on the Church of the Epiphany, Coon Rapids, MN, shown below in Fig. 1. Built in 1982, its floor plan is square with the primary structural members spanning diagonally, connecting the apse in one corner to the narthex opposite. The diagonal members serve as a cost-effective method to span 46 meters and embrace the people in the form of a gently-vaulted crucifix.

Acoustically, the space was designed to accommodate speech and music. Sound-absorptive finishes were chosen and placed strategically to mitigate troublesome echoes. Although these finishes improve speech intelligibility for the Liturgy of the Word, reverberance is choked for music; such compromises were typically employed to provide a cost-effective and functional soundscape for the Catholic liturgy. Sound-absorptive finishes in the church were common from that era with wood furring strips attached vertically over rigid, glass-fiber panels and 1” air space.

Measurements conducted throughout the nave reveal: 1) a space whose musical reverberance is too low for a space of such proportions, and 2) an amplified reinforcement (public address) system that provides only “fair” speech intelligibility as benchmarked by the speech transmission index and modulation transfer function, neither of which were the focus for this study. Although ambient noise was measured [2] as ASHRAE RC47 “Rumbly”, it was not incorporated in the auralizations.

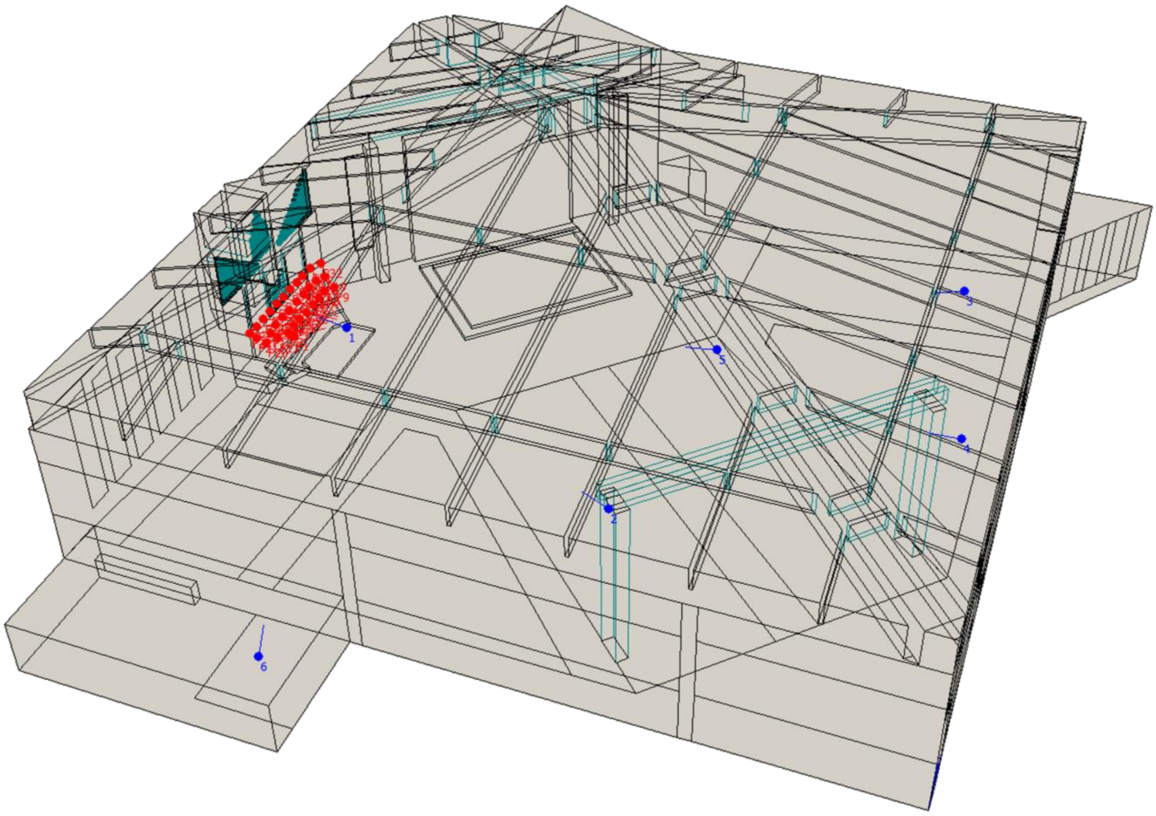

Figure 1: Isometric wireframe view of the Church of the Epiphany, Coon Rapids, MN. Choral point-source positions are shown with red dots in front of an organ facade. Receiver locations are shown with blue dots.

Architectural metrics are shown in Table 1.

Table 1: Summary of Surfaces

Surface Finish	Surface Area, m^2	Surface Area, %	Absorptivity, %
Wood T&G Ceiling	1,134	16.7	21
Wood Beams	1,957	29	11
Padded Pews	625	9.6	28
Sound Abs. Ceiling Panels	305	4.5	12
Sound Abs. Wall Panels	227	3.3	11
Carpet	471	7	7
Brick	708	10.4	…
Tile Flooring	403	6	…
Glass	333	5	…
Wood Facing	158	2.4	…
Organ Niche Wood	145	2	…
Gypsum	74	< 1	…
Organ Pipes	49	< 1	…
Doors	18	< 1	…
Total	6,805	100	100

3. Anechoic Recordings

Recordings were acquired in a standard Eckel anechoic chamber having an 80Hz cutoff frequency. Vocalists volunteered from Chesterton Academy of Edina, MN. Each vocalist was recorded individually while wearing headphones with foldback from a pre-recording of the entire ensemble and sidetones of their own voice. A conductor was also in the anechoic room with the same headphone signal conducting the vocalist; the conductor was consistently present with each vocalist during each recording. A Bruel & Kjaer 1" microphone, type 4145, was positioned one meter in front of the vocalist and one meter from the conductor; each person was on opposite corners of the anechoic room. Manuscripts were not used to prevent paper noise and comb-filtering in the recordings. An RME Fireface 800 was used to digitize, and Audacity was used to cue and sync the recordings, foldback, and sidetones. Vocal selections were chosen from the following:

1. *Sicut Cervus* — Palestrina (1525-1594)
2. *Sorida* — Rosephanye Powell (b. 1962)
3. *Give Almes of Thy Goods* — Christopher Tye (1505-1573)
4. *Hallelujah* — William Walker (1809-1875)
5. *Laudate Nomen Domini* — Christopher Tye
6. *The Pasture* — Z Randall Stroope (b. 1953)

4. Source Directivity Function

A G.R.A.S. ½" microphone, type 46AC, was systematically positioned to measure the sound radiation pattern of the mouth simulator in a G.R.A.S. 45BC KEMAR Head & Torso Simulator (HATS) in 3D and 10° resolution over the surface of a one-meter radius sphere circumscribing the HATS. Impulse responses for the directivity function were acquired at the same sampling rate and bit length as the anechoic recordings. All measurements were automated with a 3D robotic scanning system in an otherwise anechoic room.

5. Receiver Binaural HRIR

The same G.R.A.S. 45BC KEMAR HATS was measured in 3D and 10° resolution with a KEF iQ3 coaxial loudspeaker systematically positioned over the surface of a one-meter radius sphere circumscribing the HATS; the HATS contained two IEC 711 ear simulators which were calibrated and the left/right HRIR were subsequently compensated for amplitude and phase mismatch. These impulse responses were normalized by the freefield impulse response of a G.R.A.S. ½" microphone, type 46AC, placed at the center of the aforementioned sphere with the loudspeaker placed one meter in front of the manikin's nose, except with the manikin absent. All measurements were automated with a 3D robotic scanning system in an otherwise anechoic room.

6. Auralizations

The first set of auralizations were computed taking Odeon's anechoic soprano recording and convolving it at thirty-five point-source locations as shown in Fig. 1. It should be emphasized that the choir is positioned over a hard, sound-reflective floor. An auralization of the conductor's receiver position clearly revealed comb filtering from the floor reflections, which was unrealistic considering the bodies of the singers were not modelled.

In the second set of auralizations, a simple effort was made to include the sound-absorptive effect of the singers' bodies. Thirty-five point-source impulse responses were recomputed using sound-absorptive floor properties under the choral ensemble. Although this additional sound absorption prevented the unrealistic comb-filtering effect, the quality of the auralization remained synthetic in that it did not capture any realistic ensemble characteristics. Combining the same soprano from thirty-five different spatial locations didn't yield a convincing realism. This is due to the same timbre and temporal execution of the single soprano recording being used at every source position.

In both the first and second set of auralizations, an echo at the conductor position originating at the opposite wall from the choir was clearly audible, which is also evident when performing at these locations in the physical space in Coon Rapids, MN.

The third set of auralizations will be the focus of the poster session. The newly-acquired SATB recordings preserve the timbre differences of each singer and temporal variations as originally recorded/conducted in the anechoic room.

7. Computational Expense

The impulse responses for 35 point-sources and 8 receivers as shown in Fig. 1 were computed in Odeon on a Dell Workstation running Windows 7 64-bit Professional with six Intel(R) Xeon(R) X5680 CPUs at 3.33GHz and 24GB of RAM. Computation time was approximately 26 hours. Computation time for convolving 35 point-sources at one receiver position was less than 6 minutes.

REFERENCES

1 Freiheit, R., Creating an Anechoic Choral Recording, *Proceedings of the International Symposium on Room Acoustics*, Melbourne, Australia, 29-31 Aug, (2010).

2 ASHRAE Handbook of Noise and Vibration Control, *American Society of Heating Refrigerating Air-Conditioning Engineers*, Chapter 48, (2010).