This is the published version of a paper published in Journal of Voice.

Citation for the original published paper (version of record):

Journal of Voice, 1(4): 332-335

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-234454
Synthesizing Choir Singing

Sten Ternstrom, Anders Friberg, and Johan Sundberg

Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, Sweden

Summary: Analysis by synthesis is a method that has been successfully applied in many areas of scientific research. In speech research, it has proven to be an excellent tool for identifying perceptually relevant acoustical properties of sounds. This paper reports on some first attempts at synthesizing choir singing, the aim being to elucidate the importance of factors such as the frequency scatter in the fundamental and the formants. The presentation relies heavily on sound examples. Key Words: Singing—Choir—Synthesis—Chorus—Vibrato—Formants.

The sound of many voices singing the same vowel in unison is rather different from that of just one voice. We can hear that many singers are doing the same thing even when we cannot distinguish any one of them. This attribute of the multiple sound is usually called "choir" or "ensemble."

What characterizes "ensemble" sound, and how does it depend on the sounds of the individual performers? A good synthesis might answer these questions, and perhaps also shed some light on the influence of uniform vowel articulation, of different singing techniques, and of the room acoustics.

EARLIER WORK

The only detailed research we know of on synthesis of ensemble sounds is a thesis by Mark Dolson (1). He worked mostly with violin sounds, and made several interesting observations, to which we will refer to in this paper.

Several reverberation devices are commercially available, and most of these have programs for producing so-called "chorus" effects. An example is the REV7 from Yamaha, which, among other things, has two "chorus" programs. According to its manual, the effect is produced by continuously playing three copies of the original sound, each with its own delay time and amplitude which are varied slowly and out-of-phase with the other two delays. While it is more interesting than the original, the effect is not very choir-like; we perceive it as being too regular and lacking in variation. It appears to derive from the architecture of the machine rather than from the architecture of a choir sound.

ANALYSIS OF PITCH FLUCTUATIONS

In the following discussion, we will consider mainly the case of an ensemble of unison voices. Clearly, if all members of a performing group were to produce exactly the same sound, then there would be no sensation of ensemble, but of a single voice. On every sustained note, however, each singer makes small pitch and amplitude variations, even when singing without vibrato. The ensemble quality probably derives from many slightly different sounds that also are constantly changing.

What, then, is the nature of the randomness of a choir singer's pitch variations? Figure 1 shows the fundamental frequency (F) contours of choir singers singing straight tones in a choral situation. Each vertical division is one semitone. The slight tremor in F generally varies in character from singer to singer, especially with regard to the vi-
brato content. The variations are approximately plus or minus 1%. We note both rapid fluctuations, and more long-term "drift."

A convenient way of describing these F tours is to regard them as a signal and to compute its spectrum (Fig. 2). Such spectra often exhibit a peak in the 5–6 Hz region, reflecting a predominance of frequency variations at a rate of 5–6 undulations/s. This peak thus represents the vibrato.

SYNTHESIS OF VARIATIONS

In order to synthesize ensemble sounds from scratch, as it were, we first need to simulate these F contours. Even if we are ignorant of their physiological background, similar contours can be readily synthesized, using a noise generator and suitably chosen low-frequency filters. A random noise signal that is band-passed or low-passed at 6 Hz looks quite similar to the real F tour, in both the time and frequency domains, as can be seen in Figures 3 and 4.

Such simulated F contours can be used for controlling the frequency of a synthesized sound. The sound examples that follow illustrate the outcome of such syntheses. We hear a vowel tone with various "vibrato" waveforms.

These synthetic voices are rather static in character. One reason for this could be that they have no long-term pitch variation. We omitted such "drift" in these simulations, for two reasons. One was that long-term variations might indicate that the singer is searching for the ensemble mean or

FIG. 1. Examples of fundamental frequency (F contours of choir singers singing straight tones in a choral situation).

FIG. 2. Spectra of the F tours in Figure 1. Note the emphasis at 5 Hz in the dashed curve, corresponding to the vibrato.

FIG. 3. Various simulations of F tours: a) sine wave, b) noise band-passed at 6 Hz with 6 Hz bandwidth, c) the same with 2 Hz bandwidth, d) sine wave at 6 Hz plus noise low-passed at 10 Hz.
target pitch, a process which we cannot yet model. The other reason was that a random drift in each voice would lead to an unknown scatter in intonation.

MULTIPLE VOICES AND REVERBERATION

When we superimpose two identical copies of the same vowel, the sound is quite different from the single voice and exhibits a pronounced, irregular beating. The character of the sound changes if we tune the two voices apart. In the next example, the separation in fundamental frequency increases from zero to one Hz. Adding 2.6 s of reverberation also makes a difference, though not very large, for this rather stationary sound.

The question of room acoustics, or reverberation, is central to the synthesis of ensemble sounds. Acoustically, reverberation and ensemble sounds are two very similar things, both amounting to a superposition of many similar sounds. In fact, as Dolson (1) observed, it is not clear how the ear can distinguish between them. With our artificial reverberation device, we found that a single, unsteady voice, immersed in abundant reverberation, does sound very much like multiple voices. A more realistic choir-like sound results if we build a chord. We can now play a small choir of five voices, which are added one at a time. Reverberation is then added to the total.

ENSEMBLE SIZE AND VOWEL QUALITY

How does the size of ensemble affect the sound? There is a clear difference between the sound one, two or perhaps three voices, but hardly any between the sound of three and that of more voices. This agrees with Dolson's findings. We may also add several voices that are slightly out-of-tune, perhaps 1 Hz apart from each other. If we increase the spacing between voices to 2 Hz, we get a scatter of 1.5%, which sounds quite dissonant.

Choir leaders frequently wonder about the importance of having all singers pronounce a vowel in the same way. Will differences in vowel articulation affect the "ensemble" sound? Another point of interest is whether the so-called singer's formant is appropriate in choirs. Generally, scatter in vowel timbres is found to affect the choir timbre somewhat, but the ensemble quality scarcely at all.

It is possible to imitate a specific choir sound. Analysis of a recording of a unison male choir singing an "O" at about 110 Hz yielded the spectrum in Fig. 5. We can synthesize a vowel with a similar spectrum envelope, and play it five times overdubbed, with a scatter of 1% and a fair amount of reverberation. In hearing the real choir and the synthetic one alternating, the difference is clear, but not too great.

EQUIPMENT

Syntheses can be done with signal-processing software on big computers, which is slow work, or in real time with a custom-designed synthesizer. The machine we use is called SISYFOS, built by authors Anders Friberg and Kasper Marklund for their Master's theses. It consists basically of a large waveform memory for storing digitized sound and F contours, and a special signal processor that plays back the stored sounds at speeds that are pre-
cisely controlled by the stored F contours. Up to six separate voices can be played simultaneously, each with its own sound, F tour, and output channel. We are currently implementing a MIDI interface in the machine. It is intended for choir synthesis, but also for polyphonic synthesis-by-rules.

We believe that our approach is appropriate for generating effective vocal ensemble simulations in the future, thereby uncovering important aspects of the sound of choirs and of voice usage in choir singing.

Acknowledgments: This project is supported by the Swedish Council for Research in the Humanities and Social Sciences. We thank Lennart Fahlen for his continuing consultantship.

REFERENCE