Comparison of Synthesis Approaches for a Hybrid 3-Voiced Styrian Harmonica

Stephan Hohensinner[1], Matthias Frank[1]

[1] *Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria*

Email: s.hohensinner@kug.ac.at, frank@iem.at

Introduction

The Styrian Harmonica classifies as a diatonic button accordion that uses three reeds per pitch. For each pitch, these reeds are tuned to slightly different frequencies, resulting in an acoustic beat, also known as tremolo. Tremolo-related attributes have been found to be among the most relevant sound characteristics of a Styrian Harmonica. However, fine-tuning these reeds to different frequencies requires manually removing material from the reed tip. This is a very costly and impractical procedure for real instruments. Additionally, because of the expanding musical literature for the Styrian Harmonica, a single instrument cannot cover the full tonal variety.

To address these challenges, the project *Augmented-Quetschn* aims to develop a hybrid instrument that uses a single physical reed per pitch while synthesizing the other two reeds electronically. This method of sound production allows for adjustments to the perceived tremolo by controlling the amount of detuning of the high and low voices, as well as the overall timbre. Moreover, by removing two of the three reeds per pitch, the weight of the instrument is significantly reduced, making it more accessible for younger students with less physical strength to learn and play the instrument. This work explores synthesis approaches for hybrid 3-voiced Styrian Harmonica sounds within the proposed instrument design.

First, the results of a previous study are briefly discussed, including a model for the detuning of the outer reeds. The next section describes the analysis of harmonica sounds based on recordings. Then, the general synthesis approach is explained and the synthesis model for hybrid Styrian Harmonica sounds is introduced. Finally, the perceptual differences between the synthesized sounds and actual recordings are evaluated through a listening experiment, and the results are discussed in the subsequent section.

Previous Work

The analysis of harmonica sounds in this study is based on recordings originally collected for [1], which were obtained to investigate perceptual differences in artificially altered tremolos. For this, all 3-voiced tones on the melody side of a Styrian Harmonica were recorded at three manually controlled dynamic levels, corresponding to the lowest possible level (*pp*), the typical level (*mf*) and the highest possible value (*ff*). Additionally, the authors recorded all three voices per pitch separately by sealing the outlets of the corresponding reed chambers with tape. Given these recordings, the detuning of the higher- and lower-pitched voices relative to the pitch of the respective base voice has been calculated. Figure 1 shows the result of the analysis for medium loudness levels. The results suggest a symmetrical detuning, with a decreasing absolute value for higher fundamental frequencies. The dashed line represents a mathematical approximation of the detuning magnitude as a function of frequency and serves as a foundation for the synthesis. On a double-logarithmic scale, this approximation follows a linear function, allowing for changes in the detuning of the higher- and lower-pitched voices by varying the slope.

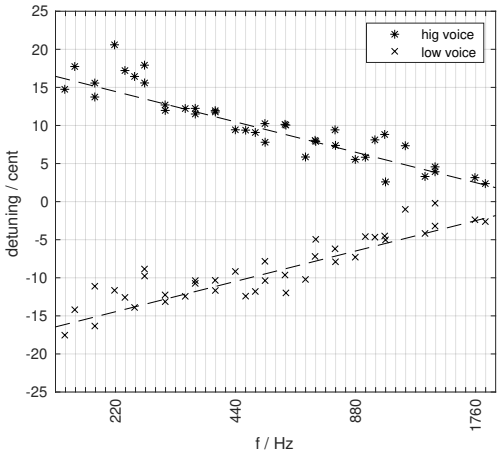


Figure 1: Figure taken from [1]: Reed detuning relative to the fundamental pitch of the base voice at a medium level.

Spectrum Analysis

The frequency spectrum of Styrian Harmonica sounds is harmonically rich and influenced by various factors, including pitch, dynamic level, the instrument itself, as well as playing style. In this analysis, the harmonic content of the lower- and higher-pitched voices is evaluated using the recorded samples played at a medium level. For this, the weights of the harmonic series – i.e., the relative amplitudes of integer multiples of the fundamental frequency – are measured. The measurements were performed using the Fourier Transform with a frequency resolution of 1 Hz. The peak amplitudes in the Fourier transform are extracted for two tones per octave across four octaves for both the higher- and lower-pitched voices up to approximately 18.5 kHz. The upper frequency limit is set to avoid aliasing in the synthesis process. To obtain general results, the harmonic weights of an octave are given by the mean of the two analyzed tones per octave, which are the notes *a* and *e*. Figure 2 shows the frequency spectra of the two analyzed notes of octave 1 with highlighted peaks, corresponding to the harmonic weights.

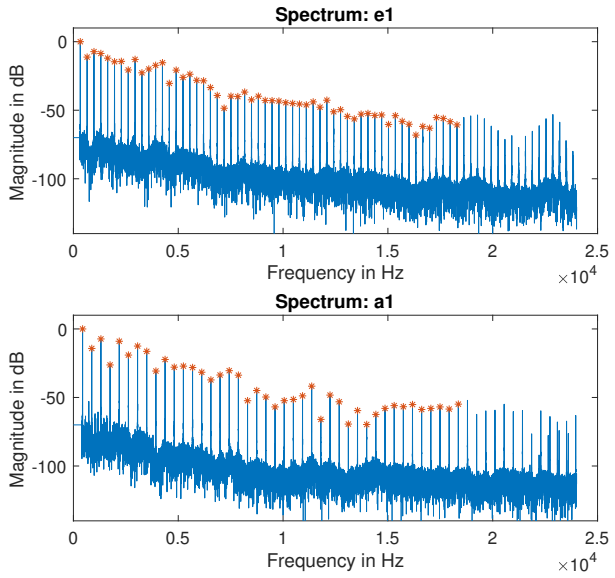


Figure 2: Normalized frequency spectra of the notes e and a in octave 1 for the low voice. Orange markers indicate the peaks, representing the amplitudes of integer multiples of the fundamental frequency up to 18.5 kHz.

Note that the number of harmonics up to the arbitrarily chosen maximum frequency depends on the fundamental frequency. Consequently, the note a does not contribute to the harmonic weights in the higher frequency range, and the number of harmonics is determined by the note e. This analysis yields the amplitudes of the first N harmonics across four octaves, with N depending on the octave. Table 1 shows the number of harmonic weights obtained per octave.

Table 1: Number of harmonics N per octave.

Octave	Number of harmonics
0	113
1	56
2	28
3	14

Waveshaping

We employed waveshaping as the method of synthesis. Waveshaping is the process of applying a transfer function, also called a waveshaping function, to an input signal. This function alters the waveform of typically simple input signals, such as sine or sawtooth waves, producing sounds with more complex harmonic structures. Figure 3 is adapted from [2] and shows the general flow chart of a waveshaper. The input signal $x(t)$ is passed to the waveshaper $w(x)$, yielding the output signal $y(t)$. The harmonic structure of the output signal depends on the choice of the waveshaping function. Typically, the waveshaper $w(x)$ is nonlinear and is represented as a polynomial, where the order determines the maximum harmonic in the resulting spectrum when using a unit-amplitude sinusoid as the input signal. For example, a linear function does not affect the resulting spectrum. However, varying the slope can attenuate or invert the input signal. Odd-order functions only produce odd-numbered multiples of the fundamental frequency, while even-order functions produce a spectrum with even-numbered harmonics. If the waveshaping function is expressed as a sum of multiple terms, the output signal will correspond to the sum of the contributions from each individual term.

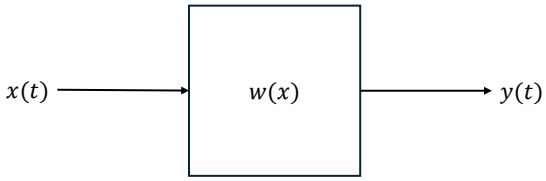


Figure 3: General signal flow of a waveshaper. The waveshaping function $w(x)$ is a function of the input signal $x(t)$.

The goal is to find a waveshaping function that produces a desired output spectrum. This method is called *spectrum matching* [2] and allows for generating sounds with complex harmonic structures by only implementing a single oscillator. The basis for such waveshaping functions is the Chebyshev polynomials of the first kind, which are recursively defined as:

$$T_{k+1}(x) = 2x \cdot T_k(x) - T_{k-1}(x), \tag{1}$$

where k denotes the order and the zeroth- and first-order polynomials are defined as:

$$T_0(x) = 1, \tag{2}$$
$$T_1(x) = x. \tag{3}$$

A common representation of Chebyshev polynomials of the first kind is expressed as:

$$T_k(\cos\theta) = \cos k\theta. \tag{4}$$

This representation exemplifies how Chebyshev polynomials can be used as waveshaping functions: Using a unit-amplitude sinusoid with the fundamental frequency θ as the input signal, the k^{th}-order Chebyshev polynomial generates an output sinusoid with a frequency that is k times the fundamental frequency. Other than single function terms, Chebyshev polynomials only produce a single harmonic. To design a waveshaping function that produces multiple harmonics with specified amplitudes, Chebyshev polynomials of different orders can be combined into a single transfer function. Consider h_k as the amplitude of the k^{th} harmonic and N as the highest harmonic in the spectrum, the transfer function can be expressed as:

$$w(x) = h_0T_0(x) + h_1T_1(x) + \cdots + h_NT_N(x) \tag{5}$$
$$= \sum_{k=0}^{N} h_kT_k(x). \tag{6}$$

Substituting the input signal x with a unit-amplitude sinusoid and using Equation 4 yields the following output

signal:

$$y = w(\cos\theta) = h_0 + h_1 \cos\theta + \cdots + h_N \cos N\theta \quad (7)$$

$$= \sum_{k=0}^{N} h_k \cos k\theta. \quad (8)$$

Note that the weight h_0, corresponding to the term $\cos 0$, represents a constant offset and is not part of the harmonic series.

Synthesis Model

For the Styrian Harmonica, four waveshaping functions were derived for both the higher- and lower-pitched voices, each corresponding to an octave within the typical range. Figure 4 illustrates the waveshaping functions, evaluated in the range of -1 to 1 with a resolution of 4800 samples. These functions are of order N, where N represents the number of analyzed harmonics per octave, and they reproduce the harmonic content of the recorded samples when a unit-amplitude sinusoid is used as the input signal. The frequencies of the input sinusoids for the lower- and higher-pitched voices correspond to the fundamental frequency of the base voice, which is detuned according to the detuning model.

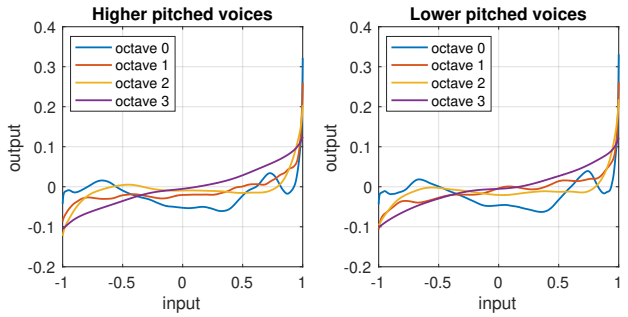


Figure 4: Waveshaping functions for lower- and higher-pitched voices obtained from harmonic weights using *spectrum matching*. These functions are evaluated in the range of -1 to 1 with a resolution of 4800 samples.

Different waveshaping functions produce a different number of harmonics. To ensure a harmonically rich spectrum across all tones and also a smooth transition between the timbres of different octaves, the outputs of the waveshapers are linearly interpolated. For this, the note g is set to be the middle of every octave, where only the waveshaping function corresponding to that octave is used. In between octaves, the outputs of the two nearest waveshaping functions are linearly interpolated according to the interpolation weights shown in Figure 5. For higher and lower tones, only the first and last waveshaping functions are used respectively.

Since the maximum frequency in the resulting spectrum does not only depend on the order of the waveshaping function but also on the frequency of the input signal, aliasing may become an issue. The maximum frequency of 18.5 kHz applies only to the analyzed tones. Tones with higher fundamental frequencies that are synthesized using this model may contain frequency components exceeding the Nyquist frequency $f_s/2 = 24$ kHz. However, additional frequency components introduced due to aliasing were not perceptible to the authors because of masking effects and the low interpolation weight assigned to higher pitches within an octave. In general, a trade-off must be made between achieving a harmonically rich spectrum across all tones and minimizing aliasing artifacts.

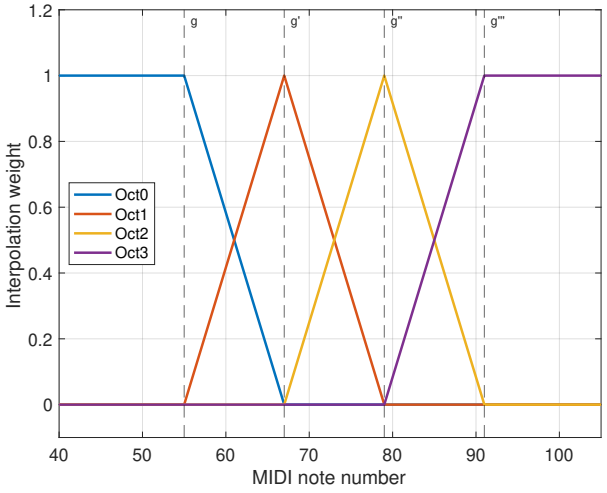
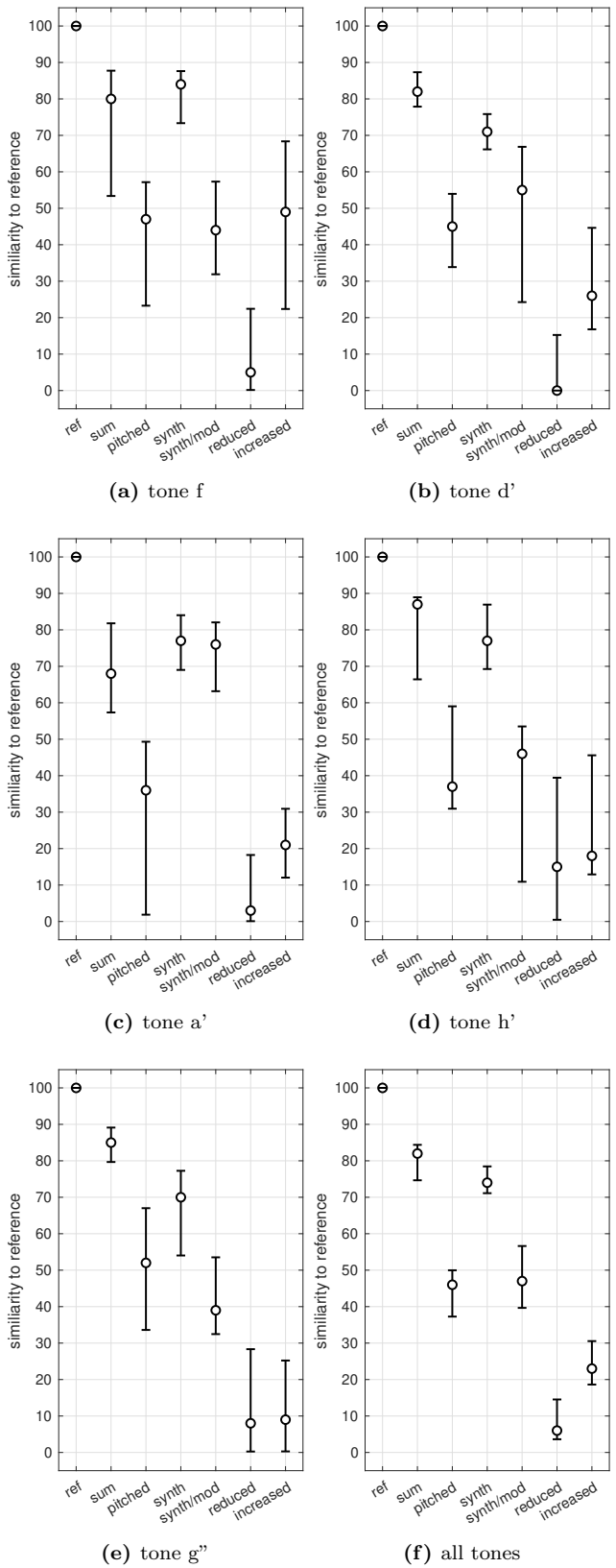


Figure 5: Linear interpolation weights for the outputs of the waveshapers, with each color representing an octave. At any given time, only two waveshaping functions are applied simultaneously.

Listening Experiment

To evaluate the perceptual attributes of the synthesized harmonica sounds, a MUSHRA-like listening experiment was conducted. Participants rated the overall similarity of the synthesized harmonica sounds to the recorded 3-voiced sound on a scale from 0 (very different) to 100 (identical). All sounds had a duration of one second, and five different tones (*f*, *d'*, *a'*, *h'*, and *g"*) were presented at a medium dynamic level. For each tone, six different versions of synthesized harmonica sounds were presented alongside a hidden reference signal (ref). The first sound corresponds to the sum of the separately recorded voices (sum). Next, the three-voiced sound is constructed by taking the recording of the base voice and applying pitch shifting upward and downward according to the detuning model to generate the high and low voice signals (pitched). We synthesized two versions of harmonica sounds using waveshaping, both based on the recording of the base voice but employing different sets of waveshaping functions. In the first version, the waveshaping function is specifically designed for the presented tones, with harmonic weights obtained from the recording of each respective tone (synth). The second version utilizes the waveshaping function from the synthesis model with generalized harmonic weights (synth/mod). For reference, two versions of the sound with artificially altered tremolo were also presented. This was achieved by pitch-shifting the recordings of the low and high voices such that the detuning was reduced to half and increased to double the amount specified by the detuning model.



(a) tone f **(b)** tone d'

(c) tone a' **(d)** tone h'

(e) tone g" **(f)** all tones

Figure 6: Perceived similarity to the reference for different tones and conditions. Each subplot represents a specific tone (f, d', a', h', g") and an aggregate across all tones. The y-axis indicates the perceived similarity to the reference stimulus (0 = very different, 100 = identical), while the x-axis displays different conditions. Circles represent the median similarity ratings, and error bars denote the 95% confidence intervals.

Results

A total of 13 participants took part in the listening experiment. Figure 6 presents the resulting medians and 95% confidence intervals for each specific tone and the aggregate of all tones. All participants successfully identified the hidden reference signal, confirming that listeners could reliably distinguish the original recording. When considering the results for all tones combined, the sum of the separately recorded voices received the highest similarity ratings among all presented harmonica sounds. However, listeners still perceived a difference compared to the actual recording of all voices, suggesting an interaction between the reeds. The synthesized sounds generated with specific waveshaping functions received similar ratings, indicating that the general synthesis approach is effective in reproducing key tonal characteristics. The sounds produced by the synthesis model were rated as intermediate between 'very different' and 'identical,' approximately matching the ratings of the pitched recording of the base voice. Obviously, the perceived similarity of sounds generated by the synthesis model strongly depends on how the waveshaping functions were derived. As an example, the similarity for the tone *a'*, cf. Figure 6c, was found to be comparable to that of the version using waveshaping functions specifically designed for this tone. The versions with altered tremolo were rated as the most dissimilar across all tones, which is consistent with expectations, as tremolo-related attributes seem to have the most significant influence on the overall sound of a Styrian Harmonica.

Conclusion

This work explored sound augmentation possibilities of a hybrid 3-voiced Styrian Harmonica. To achieve this, we developed a synthesis model that employs waveshaping to generate the sounds of the low and high voices while allowing for adjustable key sound characteristics. In a listening experiment, we evaluated the perceptual differences between synthesized and recorded samples. The results suggest that the general synthesis method produces only minor perceptual differences. However, the perceived similarity of the synthesized sounds depends on the amount of tonal information present in a given tone interval. Future refinements could include incorporating more waveshaping functions within a smaller interval to enhance similarity.

Acknowledgements

This work is supported by the project *Augmented-Quetschn*, which is funded by the Austrian Research Promotion Agency within the framework of the COIN program.

References

[1] Pomberger, H. et al.: Effect of Reed Detuning on the Sound of a 3-Voiced Styrian Harmonica. DAGA 2024 Hannover

[2] Smyth, T.: Music 270a: Waveshaping Synthesis. Department of Music, University of California, San Diego (UCSD), 2019