DOES AUDITORY MASKING EXPLAIN THE HIGH VOICE SUPERIORITY?

Song Hui Chon, David Huron

School of Music
Ohio State University

ABSTRACT

The majority of music in the world employs multiple concurrent parts. Among these parts, the upper-most part or voice often carries the melody. Considering the upward spread of auditory masking patterns, one might think the high-voice melodies are conflicting with masking theory. In this paper, we investigate the mutual masking effects of concurrent high- and low-pitched complex tones. In addition, we consider four types of spectral envelope patterns and discuss their influence on auditory masking.

1. INTRODUCTION

One of the most characteristic differences between music and speech is the typical number of concurrent sound sources involved. Speech usually involves social turn-taking, with a single speech stream alternating between the conversants. By contrast, although there is significant music-making involving a single stream (what musicians call “monophony”), the majority of the world's music-making involves multiple concurrent sound sources, and multiple concurrent auditory streams.

Among these concurrent parts, the melody, which is in general more important than the rest, is usually placed on the high voice or part. This high-voice melody practice is consistent with the fact that the changes are most easily detected in the highest stream (Zenatti, 1969; Palmer & Holleran, 1994; Crawley et al., 2002). This is known as the “High Voice Superiority.” The high voice superiority has been examined quite extensively by Trainor and her team (Fujioka et al., 2005, 2008; Marie & Trainor, 2013) who reported that it might “result from the neurophysiological characteristics of the peripheral auditory system.”

Auditory masking (ANSI, 1960) is defined as “1. The process by which the threshold of audibility for one sound is raised by the presence of another (masking) sound,” and “2. The amount by which the threshold of audibility for one sound is raised by the presence of another (masking) sound. The unit customarily used is the decibel.” In other words, masking means that when two sounds are present at the same time, one sound may not be heard as well as it would have been in isolation, due to the existence of the other sound. There are two types of auditory masking, temporal masking and frequency masking, according to the domain of the proximity of two sounds. In this paper, we will only consider the frequency masking.

Frequency masking has been studied well in the context of auditory filter and critical bandwidth (Fletcher, 1940; Greenwood, 1961; Plomp & Levelt, 1965; Scharf 1970; Patterson 1976; Moore & Glasberg, 1983; Zwicker & Fastl, 1990). Critical band usually refers to a frequency range, within which two pure tones will interact and not resolve perfectly. The masking pattern, known as the spreading function or pattern, has been obtained by examining the degree of masking effect according to the placement of two pure tones along the length of a critical band. The masking pattern is not symmetric around the center of the critical band. Rather, it has a longer tail towards the higher frequency, which is known to be the “upward spread of masking.”

This upward spread of masking means that the masking effect of a lower tone on a higher tone within the same critical bandwidth is greater than vice versa. The fact that a lower tone masks a higher tone more effectively seems to be contradictory with the High Voice Superiority, the fact that most melodies are on the high voice. Or do they really conflict with each other?

To answer this question, we implemented a computer simulation in MATLAB to examine pairs of complex tones and to determine which one masks the other better. We considered four different amplitude patterns to test their impact on masking effectiveness. Our hypothesis is that on average, higher-pitched harmonic complex tones will tend to mask the partials of comparable lower-pitched tones more than vice versa.

2. COMPUTATIONAL MODEL

A simulation, based on the psychoacoustic model in Bosi and Goldberg (2003), was implemented in order to compare the mutual masking between pairs of complex tones differing in pitch. The amount of masking will obviously depend on many factors, including the spectral envelope of the participating tones. Hence, we considered four types of spectral envelopes (“uniform,” “increasing,” “grand average” and “per-pitch average”; the last two are from Plazak et al., 2010). For each of the four envelopes, we examined their masking impacts on 81 pitches from B0 to G7. Stimuli were generated in MATLAB using sinusoids at the fundamental frequency plus 323 overtones for the first three envelope types. The sound files from Plazak et al. (2010) were used for the “per-pitch average” type.

In our simulation, we paired all possible non-unison tones spanning the range B0 to C7. That is, we began by generating complex tones for a pair of B0 with C1, each tone exhibiting the particular amplitude envelope pattern for each study. The complex tones were then used as input to a masking model. The masking model begins by carrying out a Fast Fourier Transform (FFT). The frequency and amplitude resolution of the Fourier analysis depends to some degree on whether the harmonics are integrally related to the FFT size. The FFT size we used was set to a resolution of 1 Hz, so in order to maintain good resolution we rounded the synthesized input frequencies so that they would be at one hertz integers. Hence, for example, a fundamental of 261.6 Hz was rounded to 262 Hz.

For each of paired complex tones, we calculated the masking pattern using the Terhardt spreading pattern (Terhardt, 1979). The masking envelopes for each of the lower and higher tones were individually computed. We then superimposed the masking envelope from the lower tone on the spectrum for the upper tone, and calculated the total residual power above the masked threshold. Similarly, we superimposed the masking envelope from the higher tone on the spectrum for the lower tone, and again calculated the total residual power above the masked threshold. Comparing the two total residual power values, the tone exhibiting the greater residual power was deemed to be lesser masked of the two tones. Hereafter we will refer to the lesser masked of the two tones as the "predominant" tone, and the greater masked tone as the "obscured" tone.

2.1 Study 1: Grand Average Pattern

In the first instance, we decided to limit ourselves to harmonic complex tones since these tones are most widely used in music. Musical tones exhibit highly variable spectral recipes. Even within a single musical instrument, the harmonic spectrum can differ significantly from one tone to the next and according to the method of sound generation.

Plazak, Huron and Williams (2010) calculated an average harmonic spectrum for 1,338 recorded musical instrument tones in the McGill University Master Samples (Opolko & Wapnick, 1987). The recordings included 23 common Western orchestral instruments, including piccolo, flute, oboe, English horn, clarinet, bassoon, contrabassoon, trumpet, trombone, tuba, French horn, violin, viola, cello, and contrabass. Sandell (1991) calculated harmonic spectra for each of these recorded tones. The amplitudes for each harmonic represent averages over the duration of the tone so time-variant fluctuations were ignored. The analyses were normalized so that the most energetic partial in each tone was defined as 0 dB. Plazak and colleagues calculated a grand-average harmonic spectrum for all 1,338 tones. We obtained the numerical amplitude data from Plazak for the synthesis of complex tones based on the grand average.

2.2 Study 2: Per-Pitch Average Pattern

One might reasonably object that, while the grand-average tone is representative of harmonic complex tones in general, in reality tones may exhibit patterns of spectral change with respect to register. The harmonic spectra of low-pitched instrumental tones differ somewhat from the comparable spectra for higher-pitched tones.

Accordingly, for Study 2, we made use of an "average instrument" as calculated by Plazak et al. (2010). In this case, Plazak and colleagues calculated an average harmonic spectrum for all instruments playing a given pitch. For example, the harmonic spectrum for C2 consists of the average spectrum for 12 sounds in the database. Similarly, the harmonic spectrum for C4 consists of the average spectrum for 34 sounds. In this way, Plazak et al. calculated the "per-pitch" average spectra for tones ranging from B0 to C7.

The synthesized versions of the average harmonic spectra from Plazak et al. (2010) are permanently archived in wav files at the Ohio State University's Knowledge Bank website (https://kb.osu.edu/), from where we downloaded the files. Since they were readily in the wav format, the files were read in MATLAB and the FFT results were inputted to the masking calculation model.

2.3 Study 3: Uniform Pattern

Both Studies 1 and 2 used approximations of real musical instrument tones. In both studies, the envelopes of all sounds exhibit a spectral roll-off in high frequency, which is found in many natural sounds. We then decided a priori to examine rather artificial cases, because we suspected that the spectral roll-off might contribute to the higher-pitched tone's masking of the lower-pitched tone. The first artificial pattern we considered is a uniform envelope, where all harmonics would have the same amplitude.

The stimuli were synthesized in the exactly same way as in Study 1, except that the amplitude was constant for all partials.

2.4 Study 4: Increasing Pattern

With the motivation of examining artificial envelope patterns and their impact on masking, we decided to examine another pattern that probably would not exist in nature: the increasing amplitude envelope. The amplitudes of the 324 partials were linearly increasing from 0.2 to 0.8, the values of which were arbitrarily determined a priori. In other words, for each complex tone, the amplitude of the fundamental frequency would always be 0.2 and the amplitude of the 323rd (highest) overtone would always be 0.8.

The stimuli were synthesized in the exactly same way as in Study 1, except that the amplitude was in a linearly increasing pattern.

3. RESULTS

The simulation results are summarized in Tables 1 and 2. Table 1 shows the number of occasions when the higher tone or the lower tone of a pair was predominant. The numbers in parentheses are the percentage of the total numbers of such occurrences in each category. In each of the four studies, there were a total of 3240 paired comparisons.

In Table 1, the Studies 1, 2 and 4 all have more cases of the higher predominant tone, which is consistent with our hypothesis. Study 3 with the uniform envelope pattern showed the opposite case with a more probability of a lower predominant tone. This "Low Voice Superiority" may stem from the uniform amplitude for all partials, since all other three studies have an uneven (either increasing or decreasing) amplitude patterns in frequency spectra.

The numbers in Table 1 seem to suggest that for each study, there's a tendency of either the high voice superiority or the low voice superiority. However, the proportion of the high/low voice superiority (53% to 63%) appears to be not too different from 50%, which is a random decision based on equal probability. To examine that these results are statistically significant, meaning that they are different from a random determination based on a 50/50 chance, we calculated Chi-square values and the significant values (*p*-values), which are presented in Table 2. All four studies show significant Chi-square statistics, suggesting that the numbers in Table 1 were not generated by a random 50/50 decision.

Table 1: Summary of four studies in terms of the total number of occurrences of the higher tone masking the lower tone of each pair and vice versa. In parentheses are the percentages of the total numbers of occurrences.

Study	High predominant	Low predominant
Study 1: Grand Average	**1732 (53%)**	1508 (47%)
Study 2: Per-Pitch Average	**1793 (55%)**	1447 (45%)
Study 3: Uniform	1332 (41%)	**1907 (59%)**
Study 4: Increasing	**2030 (63%)**	1210 (37%)

Table 2: Chi-square test results of the four studies to compare the results in Table 1 with a random determination based on a 50/50 chance.

Study	χ^2	p
Study 1: Grand Average	7.7525	**.0054**
Study 2: Per-Pitch Average	18.5275	**<.00005**
Study 3: Uniform	51.6078	**< 10^(-12)**
Study 4: Increasing	105.4541	**< 10^(-12)**

4. DISCUSSION

In this paper, we simulated the masking effect of pairs of non-unison tones synthesized with four different envelope types. Using tones synthesized with more realistic recipes (Studies 1 and 2), the hypothesized high voice superiority was observed. Tones with increasing amplitude patterns (Study 4), which is one of the two artificial patterns we considered, also showed the high voice superiority. The other artificial case of uniform amplitude patterns (Study 3) generated tones that exhibited the low voice superiority, against our hypothesis. This opposite behavior observed in Study 3 from all the other studies might suggest that an uneven amplitude pattern may be the fundamental reason of the high voice superiority. We will consider in future other amplitude patterns to examine this conjecture.

Overall, the results can be interpreted that placing melodies on high voices do make sense in terms of the upward spread of auditory masking, especially considering more realistic envelope patterns used in Studies 1 and 2. Since we have considered only synthesized timbres so far, it might be worthwhile to include recorded samples of real instrument sounds and repeat the analysis.

So far, we did not inspect possible patterns in terms of pitch relations. Are there some effects of register, such that a higher tone within the same octave is a predominant masker, but if the pitch difference increases it becomes an obscured tone? This investigation may reveal some hidden masking patterns based on pitch and register, which composers may have implicitly known for ages.

This paper brings the high voice superiority to a new perspective. Trainor and the colleagues have conjectured that the high voice superiority may result from auditory periphery, although they could find the effect in the auditory cortex. Our results suggest that it might be a consequence of masking in cochlea of complex tones with uneven amplitude patterns. Cochlea is at an earlier and lower level than the auditory cortex in the auditory system. Hence, our findings might be what Trainor and the team has speculated. In summary, the high-voice melodies may have been, after all, a natural development based on the upward spread of auditory masking, which we are all born with.

5. REFERENCES

American Standards Association (1960). *Acoustic Terminology SI, 1-1960*. American Standards Association, New York.

Bosi, M., & Goldberg, R. E. (2003). *Introduction to digital audio coding and standards*. Springer.

Crawley, E.J., Acker-Mills, B.E., Pastore, R.E., & Weil, S., (2002). Change detection in multi-voice music: the role of musical structure, musical training, and task demands. *J. Exp. Psychol. Hum. Percept. Perform*. *28*, 367-378.

Fletcher, H. (1940). Auditory patterns. *Reviews of modern physics*, *12*(1), 47-55.

Fujioka, T., Trainor, L.J., & Ross, B. (2008). Simultaneous pitches are encoded separately in auditory cortex: an MMNm study. Neuroreport 19, 361-366.

Fujioka, T., Trainor, L.J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J. Cogn. Neurosci*. *17*, 1578-1592.

Greenwood, D. (1961). Auditory Masking and the Critical Band, *J. Acoust. Soc. Am.*, *33*, 484-501.

Marie, C. & Trainor, L. (2013). Development of simultaneous pitch encoding: infants show a high voice superiority effect. *Cereb. Cortex*, *23*, 660e669.

Moore, B. C., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory filter bandwidths and excitation patterns. *J. Acoust. Soc. Am.*, *74*(3), 750-753.

Opolko, F. & Wapnick, J. (1987). McGill University Master Samples. (Compact discs OWCD 1-87, 2-87, 3-87). Montréal, Quebec: McGill University.

Palmer, C., & Holleran, S., 1994. Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Percept. Psychophys*. *56*, 301-312.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *J. Acoust. Soc. Am.*, *59*(3), 640-654.

Plazak, J., Huron, D., & Williams, B. (2010). Fixed Average Spectra of Orchestral Instrument Tones. *Empirical Musicology Review*, *5*(1), 10-17.

Plomp, R. & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.*, *38*, 548-560.

Sandell, G. J. (1991). A library of orchestra instrument spectra. In *Proceedings of the 1991 International Computer Music Conference*. San Francisco, CA: International Computer Music Association, 98-101.

Scharf, B. (1970). Critical bands. In *Foundations of modern auditory theory*, New York Academie, 157-202.

Terhardt, E. (1979). Calculating virtual pitch. *Hearing research*, *1*(2), 155-182.

Zenatti, A., 1969. Le développement génétique de la perception musicale. *Monogr. Francaises Psychol. 17*. CNRS, Paris.

Zwicker, E., & Fastl, H. (1990). *Psychoacoustics: Facts and Models*. Springer-Verlag, Berlin Heidelberg.