Jaatinen, Jussi; Pätynen, Jukka; Alho, Kimmo

Octave stretching phenomenon with complex tones of orchestral instruments

Published in:
Journal of the Acoustical Society of America

DOI:
10.1121/1.5131244

Published: 01/11/2019

Octave stretching phenomenon with complex tones of orchestral instruments

Jussi Jaatinen, Jukka Pätynen, and Kimmo Alho

Citation: The Journal of the Acoustical Society of America **146**, 3203 (2019); doi: 10.1121/1.5131244
View online: https://doi.org/10.1121/1.5131244
View Table of Contents: https://asa.scitation.org/toc/jas/146/5
Published by the Acoustical Society of America

Octave stretching phenomenon with complex tones of orchestral instruments

Jussi Jaatinen,[1,a)] Jukka Pätynen,[2] and Kimmo Alho[3]
[1]Faculty of Arts, Musicology, University of Helsinki, P.O. Box 24, 00014 Helsingin yliopisto, Finland
[2]Department of Computer Science, Aalto University, Konemiehentie 2, 02150 Espoo, Finland
[3]Faculty of Medicine, Psychology, University of Helsinki, P.O. Box 21, 00014 Helsingin yliopisto, Finland

(Received 13 February 2019; revised 7 October 2019; accepted 9 October 2019; published online 11 November 2019)

For decades, the phenomenon of subjectively enlarged octaves has been investigated using sinusoidal and synthesized complex tones. The present study elaborates the topic with samples of real orchestra instruments in successive tone listening experiments. Compared to previous research, this study also included a substantially larger number of subjects ($N = 36$). Examined instrument tones were categorized into five groups based on their acoustic principles. In addition, each group was assessed at three dynamic levels (pp-mf-ff). Collected data were analyzed with tuning stretch curves by applying generalized additive models in the manner of the Railsback curve used to characterize piano tuning. Although the tuning curve modeled for the orchestra instruments was observed to differ slightly from the Railsback curve and typical Steinway D grand piano tuning (Steinway, New York), the stretching trends were qualitatively similar. Deviation from a mathematical equal-tempered scale was prominent. According to statistical analyses, dynamics or musical background of the participant did not affect results significantly, but some instrument groups exhibited differences in the curve extremities. In conclusion, the stretched scale is natural for a human listener and should be used as a reference scale in tuning machines instead of the mathematical equal-tempered scale. https://doi.org/10.1121/1.5131244

[TS] Pages: 3203–3214

I. INTRODUCTION

The octave (frequency ratio of 2:1 or 1200 cents) has a special status among musical intervals in both music and research. In music, it divides the musical tone scale to the pitch chroma, which is a skeleton of the Western tonal music system.[1,2] As a scientific tool, it has certain benefits; the octave is a stable perfect consonant interval and has an equal frequency ratio in different tuning systems, such as equal temperament, Pythagorean, or just intonation. For experimental purposes, the size of an octave is rather easy for a human listener to evaluate due to the pitch class equivalence. Several psychoacoustic studies[3–14] have documented a relatively small, yet significant deviation for the enlarged subjective octave (SO) from the mathematically correct (1200 cents) physical octave (PO). Despite extensive research on this topic, earlier studies on SO have applied either simple tones as stimuli or included a relatively low number of participants. Together with a survey on previous literature, the present study explores the phenomenon with tones derived from natural instruments and a high number of participants to obtain strong ecological and statistical validity.

SO has highly important practical consequence, as the whole musical scale stretches to accommodate the enlargement from the PO. That is, the size of semitones should gradually increase as one progresses farther away from the reference tone. Thus, the larger an interval between two tones, the larger is the deviation of the interval from the mathematical tone scale (or reference scale used in traditional tuning machines). This also means that the stretched musical scale is not perfectly cyclic like the mathematical scale. The most common example of a stretched musical tone scale is present in the tuning of pianos[15] due to, at least in part, the inharmonicity of the overtones.[16] Although some type of inharmonicity is an essential feature of all piano strings, it is not significantly present with lower partials (<5). Namely, inharmonicity may explain stretch in the low-pitch region in a grand piano due to the stronger contribution of upward stretched upper partials. However, in the high-pitch region, there is no significant inharmonicity in the first four partials corresponding to one and two octaves. Nevertheless, the scale stretches similarly.[17] Sundberg[18] argues that common scale stretching and stretched tuning due to inharmonicity in pianos are unwittingly reminiscent. In piano tuning, there could be two parallel methods in use simultaneously. Piano tuning is based mainly on listening of beats in dyads (intervals), but the tuned scale should also be melodically acceptable. In the highest and lowest ends of the scale, tuners also listen to the melodic intervals as comparison, which may differ from correct beating-based tuning. In these cases, the melodic (and at the same time musical) well-tuned interval mostly overrides the beating-based (mathematical) interval. Without exception, some kind of a scale stretching exists in every concert grand piano, while inharmonicity differs appreciably among them.

The stretched musical scale is a crucial part of music perception and musical experience. However, its existence is mostly unknown due to a lack of knowledge. For a musically

[a)]Electronic mail: jussi.o.jaatinen@helsinki.fi

aware listener able to distinguish pitch with average precision, unstretched (i.e., mathematically tuned) music may sound out of tune. An experienced player or singer may use stretched tuning unconsciously.[19–21] In addition to soloists, this effect applies naturally to ensembles as well. The first author (J.J.) has an extensive career of over 30 years of professional oboe playing in several orchestras. According to subjective long-term observations, in a symphony orchestra it is mandatory to play notes in a high-pitch region higher than in the mathematical tone scale. The same phenomenon can be seen in the low-pitch region with downward tuning adjustment to avoid sounding out of tune. As a practical rule, the higher the note, the more stretching upward is needed. A tuning reference of A4 (440 Hz, or 442 Hz in this study) divides the scale into an upper and lower half. In practice, all notes below A4 should be lower than in a mathematical scale and, vice versa, all notes above A4 should be higher. This discovery was also reported by Terhardt.[22] The stretched scale can be seen as the lowest layer in the intonation system. Above that are harmonic and melodic intonation layers, similar to temperaments in the piano, as well as the intonation layer for desired artistic character. While the stretched scale layer is mainly constant, the other intonation layers are highly dependent on musical context.

The actual origin of the octave enlargement phenomenon for harmonic spectra is not well known in neurophysiology, although there is a general consensus on the mechanisms of pitch perception with the timing and place theory.[23–25] Three different theories have been presented to explain this phenomenon. One is a model by Terhardt,[26] which is based on a cochlear place theory of pitch perception and an excitation-pattern-learning model of memory, or the place theory of pitch perception, where the place refers to a tonotopical location of each frequency in the ascending auditory pathway and the auditory cortices (for additional information, see Refs. 8 and 26). Another model, proposed by Ohgushi,[27] is founded on the timing theory of pitch perception. Terhardt's[26] central template theory is based on mutual masking of the upper tone. In contrast, Ohgushi argues that the refraction period of the peripheral auditory nerves produces a shift to the subharmonic interspike intervals (see also Ref. 28). McKinney and Delgutte[29] have developed Ohgushi's model further and improved its reliability. The third and newest explanation has been presented by Bell and Jedrzejczak.[30] They proposed that in the cochlea synchronized spontaneous otoacoustic emissions elicit a frequency ratio 1.063 ± 0.005, which is close to a semitone stretched by a small amount. When this ratio is raised to the power of 12, the result is an enlarged octave ratio of approximately 1:2.08. Hence, if the semitone is already stretched, the octave enlargement phenomenon could also have its roots at the cochlear level. Due to an individual variability between subjects in measured frequency ratios (i.e., stretched semitones), it also influences the perceived size of the enlarged octave.

All the aforementioned theories are based on sinusoidal stimuli and leave many details unexplained. Moreover, no recent neurophysiological studies are available. The neural basis of this phenomenon is much likely more complicated. Although the aforementioned theories may provide a partial explanation, further investigation is needed, especially with complex harmonic tones.

Although the stretching phenomenon of SO has been explored with many approaches, the existing literature lacks a thorough and commensurate study that considers several relevant musical aspects of the SO phenomenon. In present study, we report research on the stretched octave charting the effect with realistic instrument spectra, and propose a tuning curve for an orchestra setting analogous to preferred piano tuning.

II. BACKGROUND AND LITERATURE REVIEW

Pure or harmonic complex tones in intervals outside the mathematical octave tuning cause audible fluctuation in simultaneous presentations. The effect of phase-locking renders the mathematical octaves easily distinguishable from the stretched and contracted octaves, which makes unbiased comparison very challenging if not impossible (see also Ref. 12). Therefore, almost all octave enlargement studies have used a design of successive tones. In most earlier experiments, the lower tone has been used as a reference tone and the upper one has been adjustable.

In the earliest study, published by Stumpf and Meyer[3] in 1898, musically perceived consonance of different musical intervals was measured, including an octave (300–600 Hz). For producing sounds, they used an Appunscher Tonmesser [Appun reed tonometer, a box with 102 reeds that can produce tones from 128 to 1024 Hz with resolution of 4 Hz (Ref. 31)] and tuning forks as in the octave experiment. The lower tone was 300 Hz, and the upper tones varied between 598, 600, and 602 Hz. The most preferred octave was the largest alternative (602 Hz), which corresponds to a difference of circa 6 cents over the mathematical octave.

The study by von Maltzew[14] published in 1913 is based on interval recognitions. According to her findings, in a high-pitch region (range from C3 to E6), a mathematical octave was most often recognized as a major seventh. As a stimulus source, she used specially designed flue pipes.

The first experiment that focused purely on octave enlargement was published by Ward[4] in 1954. Since then, his experimental design has been adapted by the majority of later studies on the same topic. Ward was also the first to employ an electronic device as a stimulus generator with adjustable tone freely controlled by a participant. Sinusoidal tones were controlled with Fletcher-Munson equal-loudness contours to equalize presentation sound level over different frequency bands. Ward measured the left and right ears separately and the binaural diplacusis (interaural pitch difference) was also considered. Although there were intrasubject differences between the ears, the mean data (within and across participants) clearly demonstrated the octave enlargement phenomenon. The amount of stretching was observed to be highly frequency dependent and increased in higher pitch regions. However, when approaching the upper limit of the human pitch perception ability[2,34] of circa 5 kHz (here, observations above 4 kHz), deviation increased significantly.

In a study published in 1969, Walliser[5] used downward octaves, that is, the upper tone was a constant reference tone and the lower tone was adjustable. However, the obtained results were comparable to Ward's, and no significant difference between the upper or lower reference was found.

In a subsequent study from 1971, Terhardt[6] was the first one to use synthetically produced complex tones as stimuli. In comparison with previous studies using sinusoidal stimuli, the amount of octave enlargement of the complex tones in pitch regions over 2 kHz was slightly smaller (~1% or ~17 cents), while the general stretching trend remained comparable.

Sundberg and Lindqvist[7] (1973) also applied complex tones as stimuli while introducing varying sound level as a new aspect. They used three different sound pressure levels (SPLs) in reference tone (65, 80, and 90–95 dB). The amount of stretch was observed to vary with stimulus intensity and was also frequency dependent. In the lowest and highest pitch regions, octave stretching was significantly larger. In case of stimulus intensity, no simple formula was proposed for explanation. Another novel finding was the similarity of the observed scales with stretched scales recorded from real musical performances, where notes played in the high-pitch region stretched upward.

Dobbins and Cuddy[9] (1982) employed both musicians and non-musicians as participants. This study revealed that musical training did not significantly influence the amount of stretch but decreased intra- and inter-individual variability.

Ohgushi[27] (1983) assumed a neurophysiological focus in his study as mentioned earlier. He also conducted a small behavioral test with both successive and simultaneous pure-tone pairs with results similar to earlier octave enlargement findings (see also Fig. 1). Strangely, the obvious fluctuation that should occur with simultaneously presented tones was not reported, albeit measured octaves were stretched.

Demany and Semal[12] (1990) distinguished melodic and harmonic octave templates by comparing successive and simultaneous presented octaves. By requiring the listener to use a harmonic listening template and avoiding fluctuations other than those in the mathematical octave, one of the simultaneously presented pairs were frequency modulated by 2 or 4 Hz. As a result, variability in harmonic (simultaneous) adjustments was larger than in melodic (successive) adjustments and

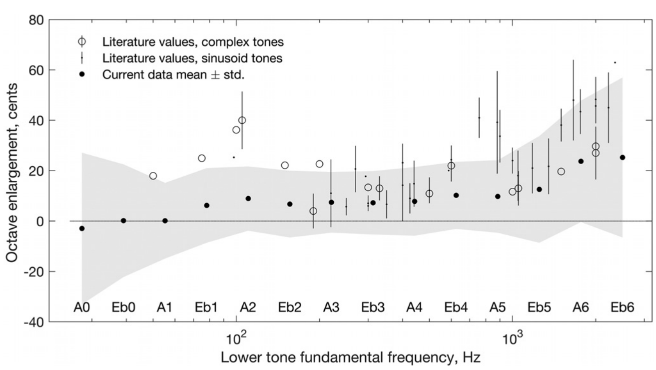

FIG. 1. Collected literature values of SO stretching and respective reported standard errors grouped according to the type of the stimulus signal. Single dots and shaded area denote the results from the current experiments as described in Sec. IV.

increased with frequency. However, due to the small number of participants (three), further investigations are needed.

Hartmann[8] (1993) compared the neural origin theories of this phenomenon (Terhardt[26] and Ohgushi[27]) and included Huggins' tone as stimulus type. Huggins' tone refers to a pitch that can only be heard if listened to binaurally. It is generated by presenting white noise to both ears; signals are in identical amplitude, but in one ear a narrow frequency region is phase-shifted 180 degrees. The shifted region can be detected as a sensation of pitch.[32] Octave enlargement phenomenon was elicited also by Huggins' tone stimuli. Since a dichotic pitch is assumed to be perceived in a higher stage of the auditory system, the result in part supports Terhardt's central templated theory. However, Hartmann did not completely reject Ohgushi's timing theory either.

Rosner[13] (1999) examined sinusoidally produced major seconds, perfect fourths, and octaves by six professional string players. The results suggested octave enlargement, whereas seconds and fourths tended to be contracted.

Previous comparable research is shown in Table I together with the stimulus type and their presentation mode and intensity and the number and type of participants.

III. METHODS

In the current study, we approached the octave enlargement phenomenon from a musical starting point. We used real tones produced with real orchestra instruments and dynamics relevant in a musical context (pp-mf-ff). For the first time, all multiple octaves available in symphony orchestra (1–7 octaves) were measured. The tuning method used in the present study enabled stimulus presentation fundamental frequency accuracy of 0.003 cents, although leaving spectral content untouched. Compared with all prior studies, the number of participants ($N = 36$) was substantially larger. In the present design, the octave enlargement phenomenon could be investigated equally with single octaves and multi-octave intervals.

Data on the octave enlargement phenomenon were gathered with a listening experiment. In the manner of previous studies, the participants listened to pairs of tones where the lower and higher tones with the same signal type in a single or multi-octave interval alternated in a 1-s period each in repeat until accepted by the participant. The task for the participants was to adjust the higher tone to their perceived SO. The tone pairs included all possible octave multiples from single octaves up to a seven-octave interval.

The adjustment input was implemented as keystrokes that switched between 41 pre-calculated tuning increments over the range of [−72,+72] cents with respect to PO. To provide a wider adjustment range without increasing the number of tunings, the increments were slightly larger outside the range of [−36,+36] cents. From 0 to 36 cents the increment was 3 cents, between 36 and 52 cents the increment was 4 cents, and beyond that the increment was 5 cents. This non-linear adjustment resolution was not disclosed to the subjects.

The stimuli represented five groups of typical orchestra instruments: flutes, single-reed woodwinds (i.e., clarinets), double-reed woodwinds (i.e., oboe and bassoons), brass, and

TABLE I. Previous research and the present study on the enlargement of the SO. *N* denotes the total number of subjects.

Study	Year	Signal[a]	Presentation[b]	Reference tone[c]	SPL	Participants[d]						Gender[d]	
						N	PM	SM	AM	NM	U	F	M
Stumpf and Meyer (Ref. 3)	1898	TF	SU	L	—	18	18	0	0	0	0	0	18
von Maltzew (Ref. 14)	1913	FP	SU	L	—	15	1	2	11	0	1	4	11
Ward (Ref. 4)	1954	SI	SU	L	~50	9	0	6	3	0	0	—	—
Walliser (Ref. 5)	1969	SI	SU	U	~60	4	0	0	4	0	0	—	—
Terhardt (Ref. 6)	1971	SC	SU	L	~60	7	0	0	0	0	7	0	7
Sundberg and Lindqvist (Ref. 7)	1973	SC	SU	L	~65–95	4	0	0	4	0	0	—	—
Dobbins and Cuddy (Ref. 9)	1982	SI	SU	L	~50	6	0	3	0	3	0	—	—
Ohgushi (Ref. 27)	1983	SI	SU, ST	L	~35	3	0	0	0	0	3	—	—
Demany and Semal (Ref. 12)	1990	SI, SC	SU, ST	L	~45	3	0	1	2	0	0	—	—
Hartmann (Ref. 8)	1993	SI, HT	SU	L	~54	6	0	0	4	0	2	2	4
Rosner (Ref. 13)	1999	SI	SU	L	~55	6	0	0	6	0	0	5	1
Present study		RC	SU	L	~65–76	36	18	0	18	0	0	7	29

[a]TF = Tuning fork, FP = flue pipes, SI = sinusoidal, SC = synthesized complex, HT = Huggins' tone, RC = real instrument complex.
[b]SU = successive, ST = simultaneous.
[c]U = upper, L = lower.
[d]PM = Full-time professional musicians, SM = full-time student in music, AM = amateur musicians, NM = non-musicians, U = unknown, F = females, M = males, — = unknown or unreported value.

strings. Instruments were grouped by their acoustic principles. The presented octaves spanned the entire ordinary compass of each instrument group, including the low- and high-pitch region versions of different instruments. The spectral variation of different playing dynamics was considered by including three nominal dynamic levels from each instrument. The instrument tones included in the final experiment configuration were A0–A6 and E♭1–E♭6 (brass), A1–A6 and E♭1–E♭7 (clarinet), A0–A6 and E♭1–E♭6 (double reeds), A3–A7 and E♭4–E♭8 (flutes), and A0–A7 and E♭1–E♭8 (strings). Since the experiments contained all octave combinations within particular instruments, A and E♭ tones and three dynamic levels yielded a total number of 552 trials. An additional 12 trials for charting the effect of small shifts in tuning reference were included in random order among the test sequence to validate the concept of the tuning curve. This part of the experiment is described in more detail in Sec. III C.

The listening experiment was organized in three identical acoustically isolated listening booths (Aalto Acoustics Lab, Espoo, Finland) with Sennheiser HD650 headphones (Wedemark, Germany) driven with USB headphone amplifiers running at 48 kHz sample rate. The keyboard and mouse were inside the booth, and the display showing the graphical experiment interface was placed behind a window. The test procedure, including the user interface and audio playback, was programmed in Max environment (Cycling'74, San Francisco, CA). The test routine prompted for a break for every 60 trials. The designated duration for the entire experiment, including the audiometry screening and breaks, was approximately 3 h.

A. Participants

Due to the inherent difficulty of distinguishing small deviations in the sizes of octaves, a musical background is advisable. The present experiments included both professional ($n = 18$) and amateur ($n = 18$) musicians as participants ($n = 36$, ages 20–64 years, mean 45.6 years, standard deviation 12.3, 7 females, 29 males). There were two representatives of all major orchestra instruments (one professional and one amateur): flute, oboe, clarinet, bassoon, violin, viola, violoncello, double bass, trumpet, horn, trombone, tuba, piano, saxophone (the amateur was the author K.A.), harp, and male vocals. In addition, there were two extra woodwind players and the two authors (J.J., oboe, and J.P., piano). Two participants (the authors J.J. and J.P.) had absolute pitch. Professional musicians were from the highest-level symphony orchestras in Finland (Finnish Radio Symphony Orchestra, Helsinki Philharmonic Orchestra, Tapiola Sinfonietta, Lahti Symphony Orchestra, Turku Philharmonic Orchestra) or worked as teachers or professors in the Sibelius-Academy, University of Arts, Helsinki, Finland. All of the amateur musicians had a long instrumental background and good knowledge in music theory and terminology. Most of them played in amateur symphony orchestras, and some graduated from the secondary music schools.

B. Stimuli

The stimuli for the listening experiments were built upon a steady-state wavetable synthesis. The signals were gathered mainly from a sample library designed for professional music production (Vienna Symphonic Library GmbH, Austria). The sample library contains sampled voices from the most common orchestra instruments recorded professionally in an acoustically controlled studio (0.8 s reverberation time) over a wide range of fundamental frequencies. Additional samples at frequency extrema were captured in a recording studio (1.0 s reverberation time) by the authors with professional musicians.

The generation of the signals consisted of the following procedure. First, the waveform of a particular tone was upsampled to the sample rate of 384 kHz, after which a single period was isolated from the waveform with WaveLab Pro 9.5 software (Steinberg GmbH, Hamburg, Germany). Second, the waveforms were imported to MATLAB environment for accurate regulation of the fundamental frequency. The fundamental frequency was adjusted to the nominal equal tempered pitch re. A4 = 442 Hz. This result was accomplished by repeating the single-period waveform with an integer count until a duration of 2 s was reached. Correct tuning was achieved by resampling the repeated waveform by a ratio of integers that was searched iteratively based on the knowledge of number of fundamental periods in the repeated waveform and the ideal duration for the desired tuning frequency. In addition to applying this procedure for the lower reference tone, this method was used to synthesize the pre-calculated tuning increments for the higher note for the subjects' adjustment. The accuracy of this regulation process (approximately 0.003 cents) greatly surpassed the resolution of adjustment shown for the subjects. After the regulation, the amplitudes of each resulting signal were equalized with *C*-weighting to the same value as the note A4 with flute in the respective playing dynamics. Finally, the signals were resampled to the presentation sample rate of 48 kHz, truncated to a length of 1 s, and onset and offset transients were tapered with a Tukey window with 35 ms fade-in and fade-out.

The equivalent levels ($L_{\mathrm{eq,A}}$) of stimuli representing different dynamic levels were 65.1, 71.2, and 75.6 dB (SPL) for pianissimo, mezzoforte, and fortissimo, respectively. For some instruments, the resulting range of dynamics was narrower than in reality, but the presented dynamic range was still feasible for listening to the stimuli for the duration of the experiment. In particular, the authors tried to avoid the annoyance in evaluating flute or trumpet tones in high registers.

C. Octave additivity property

Our aim was to estimate a tuning curve based on the tuning values obtained with different orchestra instruments. Estimating the tuning curves over the tested range of tones is based on the assumed additivity property of octave tuning, as shown earlier by Ward.[4] This assumption means that the stretching of single octaves becomes accumulated for the respective multi-octave interval, that is, $\mathrm{SO}_{\mathrm{A4\text{-}A6}} = \mathrm{SO}_{\mathrm{A4\text{-}A5}} + \mathrm{SO}_{\mathrm{A5\text{-}A6}}$. This additivity for formulating a tuning curve also requires that small shifts in fundamental frequency of the reference tone do not influence the size of the SO.

To validate this requirement, the experiment also included special tone pairs of single-octave (A3-A4) and triple-octave (A2–A5) intervals. In addition to the unaltered tone pairs, the presented trials included four versions with 5-, 25-, and 200-cent positive shifts in the lower tone, as well as in the range of the adjustable upper tone. These combinations were synthesized with mezzo forte spectra of two instrument classes (double reeds and strings). In total, the subjects produced a total of 432 additional adjustments (2 intervals × 2 instruments × 36 subjects × 3 small-shift variations). Furthermore, a 600-cent positive shift was utilized from the respective original E♭ tone pair data. Analysis on the small frequency shifts is reported separately in Sec. III D.

D. Algorithm for octave stretching over playing range

The stretching curve is calculated upon a reference point, namely, the tuning center. Being the standard pitch reference, A4 is a natural choice as the reference point. Single-octave interval tuning values above and below the center were stacked according to the additivity assumption. As all trials were produced by adjusting the higher tone, the tuning values for tones below the reference tone were inverted to always accumulate the octave stretching away from the tuning center. The process started near the tuning center and proceeded iteratively toward the low and high octave extrema according to the assumed additivity of tuning values. By repeating the procedure separately for data on each combination of participant, instrument, and dynamics, the process yielded distinct tuning curves for each independent variable.

In addition to the single octaves, the listening experiment included all possible multi-octave intervals. Where compiling the tuning curve for single octaves is straightforward, establishing a comparable curve for arbitrary multi-octaves required a more elaborate procedure. We used computation in three parts depending on the condition of the multi-octave interval in relation to the tuning center: (1) if either of the tones were equal to the tuning center, the tuning value would be attributed to the octave outside the tuning center correspondingly to the single-octave condition; (2) if the multi-octave interval crossed the tuning center, the tuning value from the trial was divided for the lower and higher tones of the interval in proportion to the tones' distances from the tuning center; (3) if the span of the interval was outside the tuning center, the tuning value was attributed to the tone more distant from the tuning center. In addition, the tuning value of the tone closer to the center from single-octave data was added, adhering again to the additivity assumption. Together, these rules produced one or more values for each tone in the range of octaves. The data points forming the stretching curve over these tones were obtained as the median over different values respective for one tone. The entire procedure was repeated separately for double, triple, or other multi-octaves for comparing the subjective stretching of the larger intervals.

Since octave multiples of one tone only give a relatively sparse sampling of the overall frequency scale, the same procedures were repeated correspondingly for both tones A and E♭, where the tuning center for the latter was E♭4. To incorporate the E♭ notes to the stretching curves of A octaves, the tuning values obtained for all E♭ notes were adjusted by a constant that was obtained as the interpolated tuning value between the adjacent values for the A notes, that is, A3 and A4 (the tuning center). This approximation assumed that the octave stretching occurs linearly within the region of interpolation, which appeared reasonably valid based on the data presented in Secs. III E and IV.

E. Statistical analysis

The nature of dependent variable data obtained from the experiment followed a discrete probability distribution as the participants were presented with a fixed set of alternatives for desired tuning. The independent variables, namely, the frequency, also all have discrete values.

A preliminary investigation of the data suggested that the distribution of raw tuning curve data points at separate references frequencies was slightly skewed in the majority of cases. This property violates the assumption of normality, which was also observed from the outcome of Lilliefors tests for normal distribution. However, the number of samples was sufficiently high to not exclude particular statistical approaches such as non-parametric methods.

A substantial part of previous studies analyzed the octave enlargement effect as a set of separately sampled data points. The obtained results have largely been reported only in numeric format for the sampled fundamental frequencies. Conversely, an overall model for the octave enlargement phenomenon has not been constructed. As suggested by the Railsback curve, the octave enlargement phenomenon could potentially be generalized with a continuous curve. An earlier study by Rosner presents an empirically fitted non-linear curve to the SO literature values.[13] In addition to the work by Rosner, Sundberg and Lindqvist[7] also modeled their octave enlargement experiment results with continuous functions, in a manner of the Railsback curve.

The present study modeled formed tuning curves through generalized additive models (GAMs).[33] In principle, GAM models the data as a combination of smooth terms. A significant advantage of GAM is its capability to model arbitrary shapes that would not be achieved by straightforward polynomial fits. High-degree polynomial models for data would also produce challenges when comparing fitted subsets of data for different instrument groups or dynamic levels. GAM analysis does not require *a priori* information or assumption of the polynomial degree, but instead aims to minimize degrees of freedom while retaining an accurate fit.

GAM analysis was performed in *R* environment with the `mgcv` package, which estimates the models by maximizing the penalized likelihood of piecewise linear functions or splines. Among several options for spline types, current data with multiple simultaneous variables and interactions within those advocate for selecting thin plate splines, which is also the default smooth term in the `mgcv` package.

Similar to conventional linear models, GAM is sensitive to the violations of homoscedasticity. For this reason, the data were weighted by the inverse of local variance per tone frequency $1/(1+\mathrm{var}(\vec{x}))$, where $\vec{x}$ is the subset of data for the respective combination of factors. Since the variance of data was higher at the frequency extrema, the above weighting by variance further widens the confidence intervals that are eventually associated by estimating the statistical differences between compared instrument groups or dynamics.

Analysis of variance (ANOVA) is available for GAMs in a similar manner as conventional linear models. Thus, ANOVA for GAM performs hypothesis tests relating to the desired GAMs. For one fitted GAM, ANOVA consists of Wald tests of significance for each parametric and smooth term, so interpretation is analogous to type III ANOVA rather than a sequential type I ANOVA. Comparison of fitted models can be performed as one-way ANOVA. However, the authors were not aware if two-way ANOVA for resolving possible underlying interaction between GAM fits is possible. Therefore, the two-way analysis was performed with the more conventional method of inspecting instrument and dynamics groups separately at the pitch regions above and below A4 with linear approximation.

IV. RESULTS

The results from the present experiment are reported in specific parts in Secs. IV A–IV F. First, we concentrate on the validation of the octave additivity assumption in the current data. Second, the enlargement of separate single octaves is presented. Finally, we present outcomes of the statistical analyses and comparisons of tuning curves over different variables.

A. Octave additivity

The additivity property is the principal assumption in forming the tuning curves from raw listening experiment data. The applicability of this property should be inspected from two views. First, the equality of subjective multi-octave and corresponding successive single SO have to produce equal stretching. This relation has been already verified by Ward.[4] Second, the additivity assumption requires that small shifts in the fundamental frequency of the reference tone do not influence the size of the SO. The sensitivity of the SO to small shifts in the reference tone's fundamental frequency has not been ensured previously. To verify this effect, we analyzed the set of single-octave or triple-octave tone pairs, which had slightly different fundamental frequency in the lower tone, as described in Sec. III.

The data were analyzed by linear model with the full interaction model *Shift**Instrument*OctaveRange* (4*2*2 levels), aiming at resolving whether any of the variables had an influence on the SO enlargement. The instrument group type did not have a significant effect on the SO ($t_{700} = -1.090$, $p = 0.28$). The overall shift of 25 cents suggested a slight deviation from the other frequency shifts as a main effect but not at a statistically significant level ($t_{700} = -1.733$, $p = 0.08$). Only the factor of octave range showed a statistically significant effect on the SO ($t_{700} = 4.332$, $p < 0.001$). This outcome is consistent with the octave additivity assumption in two aspects. First, the size of SO is not sensitive to small shifts in the reference tone frequency. Second, the SO enlargement is expected to be wider with multi-octave intervals. On these bases, the octave additivity can be assumed valid for the subsequent analyses.

B. Overall single-octave stretching

The enlargement of the SO is directly obtainable from the listening experiment data. The overall range of obtained single-octave stretching values over the experimented reference tone frequencies is demonstrated against the comparable results in

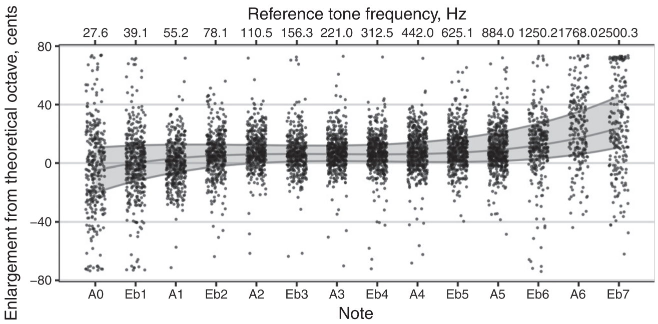

FIG. 2. Overall single-octave enlargement from the listening experiment. The curve shows an approximation of the median values with cubic quantile regression with the equation $Y = 0.44X^3 + 1.33X^2 + X + 6.22$, where X denotes the fundamental frequency of the reference tone in octaves from A4, and Y is the octave enlargement in cents. The shaded region indicates the 25th and 75th percentiles.

the previous literature in Fig. 1. The grand mean over all variables with single octaves is shown in Fig. 2. Although the variation is high, the general trend that the SO is increasingly enlarged toward high frequencies is evident. The results from the highest octaves from A6 upward show a less regular distribution of the tuning values. In particular, at the octave E♭7-E♭8 where the reference tone has a fundamental frequency of 2500 Hz (upper limit of pitch perception[34]), the consistency of the SO adjustment was vague in comparison to the middle frequencies. Furthermore, part of the data shows a saturation to the upper limit of the adjustment range. A corresponding effect is also seen at the extremely low frequencies, mainly at A0 (the lower limit of pitch perception[35]).

The statistical analysis consisted of a linear model with a main effect over the independent variables of lower tone fundamental frequency (14 levels), instrument class (5 levels), and dynamics (3 levels). The results are tabulated in Table II.

A continuous estimate of the SO enlargement is included in Fig. 2 by cubic polynomial quantile regression. The regression yields a relation $Y = 0.44X^3 + 1.33X^2 + X + 6.22$, where the lower tone X of the evaluated interval is given as octave distance from A4, and Y is the estimated difference from the PO tuning in cents. Notably, the estimate remains monotonously increasing. The quantile regression method is chosen here due to the heteroscedasticity at the frequency extrema, which suggests deviating from the traditional least-squares estimation. However, as a result of the large number of data points, the corresponding linear model produces a nearly identical estimate of the SO enlargement. Polynomial coefficients beyond the cubic model do not appear to provide statistical significance.

C. Tuning curve with single octaves

The data on octave enlargement can be presented in an alternative form that corresponds to piano tuning curves. Instead of local enlargement of SO, the tuning curve shows the cumulative stretching effect over the entire range of octaves in relation to a single reference tone (A4). An average preferred size of arbitrary octaves can be directly estimated from the curve by subtracting the y axis value of the lower note from the corresponding value of the upper note.

TABLE II. Results from a linear model for single-octave SO in cents at different fundamental frequencies, instruments, and dynamic levels. All p-values are Bonferroni corrected. Total number of samples was 5940 (***$p < 0.001$, *$p < 0.05$).

	Estimate		Standard error	t-value	p-value
Note A0	−2.650	*	1.099	−2.411	0.016
Note E♭1	0.300		0.999	0.300	0.764
Note A1	0.288		0.999	0.288	0.773
Note E♭2	6.355	***	0.999	6.359	<0.001
Note A2	9.085	***	0.999	9.089	<0.001
Note E♭3	6.876	***	0.999	6.880	<0.001
Note A3	7.480	***	0.941	7.948	<0.001
Note E♭4	7.270	***	0.941	7.726	<0.001
Note A4	7.835	***	0.941	8.326	<0.001
Note E♭5	10.228	***	0.941	10.869	<0.001
Note A5	9.787	***	0.941	10.400	<0.001
Note E♭6	12.599	***	1.193	10.556	<0.001
Note A6	23.920	***	1.403	17.049	<0.001
Note E♭7	25.443	***	1.403	18.134	<0.001
Instrument clarinet	0.319		0.727	0.439	0.661
Instrument double reeds	−0.042		0.716	−0.059	0.953
Instrument flute	0.475		0.857	0.554	0.579
Instrument strings	−0.921		0.696	−1.324	0.186
Dynamics *mf*	−0.026		0.554	−0.047	0.962
Dynamics *ff*	0.061		0.554	0.110	0.912

The grand average tuning curve derived from single-octave raw results over five instrument groups is presented in Fig. 3. The visualization includes individual data points as well as the respective tuning curves estimated with GAM without predetermined degrees of freedom and data weighting by inverse variance as described in Sec. III E. GAM estimation with Gaussian link function yields smoothing with an estimated 6.416 degrees of freedom ($p < 0.001$), which is analogous to the degree of polynomial fit. The residual effect degrees of freedom is 7.5. In terms of a conventional F-test, the rounded degrees of freedom correspond to ($F_{6,8}$). This result indicates a statistically significant difference from a flat, unstretched tuning curve. The obtained tuning curve within the range of A2–A6 follows a fairly linear relationship with the pitch, as suggested by the nearly constant single-octave enlargement in Fig. 2. In contrast, the variance of individually accumulated data points is substantial outside

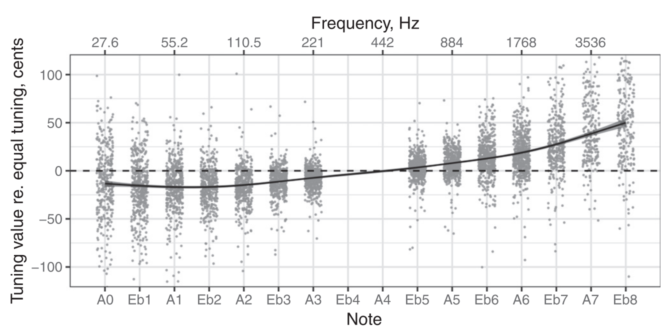

FIG. 3. Grand average tuning curves estimated with generalized additive models (GAM) for the entire dataset. Underlying datapoints are treated with horizontal and vertical jitter within the pitch and tuning adjustment resolution. The shaded region visible mostly in the frequency extrema indicate the 95% confidence interval for the values predicted by GAM.

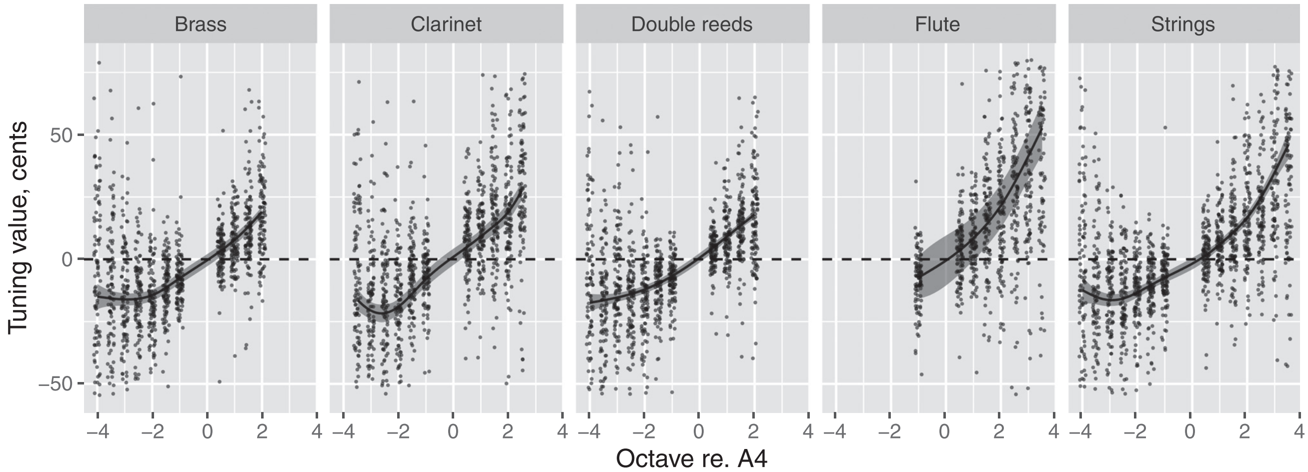

FIG. 4. Tuning curves estimated with GAM for each instrument group from the single-octave data. The overlaid ribbon indicates the estimated mean and 95% confidence intervals for each note.

the range of A2–A6, which is described by the adjusted R^2 of 0.33. In these regions, the estimated tuning curve features a slight curvature upward. Although the data variance appears large by visual inspection, the estimated confidence intervals for the fitted model remain negligible.

Analysis by instrument groups is visualized in Fig. 4. At this point we can observe that all instrument groups follow the general trend of the grand average. However, the curve with clarinets exhibits a stronger non-linearity near octave indices 1–2. The flute curve, in contrast, demonstrates larger deviation between the participants, which results in considerably wider confidence intervals for the tuning curve estimate. Residuals of fitted curves showed no apparent trend, and as can be observed from Fig. 3, the residual deviations are more consistent at the middle-pitch region than near the extrema. Tuning curves estimated for each instrument group are compared in Fig. 5, which emphasizes generally corresponding shapes. Estimated degrees of freedom vary between 2.69 for double reeds (closest to cubic fit) and 5.46 (most complex curve) for both strings and clarinets. Their mutual similarity is statistically tested with one-way ANOVA for the GAM model, where the comparison of parametric terms yield ($F_{4,26} = 1.74$, $p = 0.14$). This result suggests that, despite some apparent differences, the tuning curves of instrument groups are not statistically significantly different from each other. It is worth noting that the inverse variance weighting for the GAM analysis applies also here.

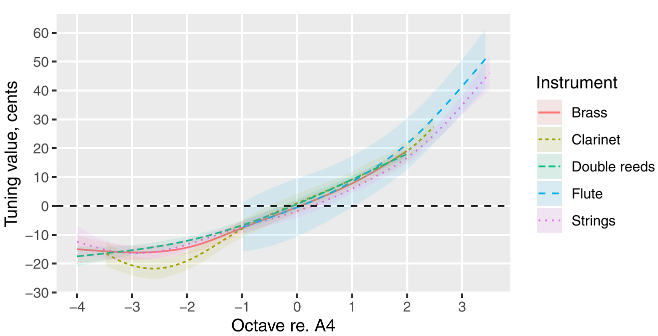

FIG. 5. (Color online) Comparison of estimated mean tuning curves with each instrument group and respective confidence intervals in shaded areas.

A similar comparison was conducted over the tuning curves respective to three overall dynamic levels averaged over instrument groups. With visual inspection, the curve for mezzoforte shows a marginally emphasized J-shaped curve like the clarinet in Fig. 5. However, the dynamics did not show statistically significant differences on the tuning curves with ($F_{2,19} = 1.063$, $p = 0.345$).

D. Difference between single- and multi-octave tuning curves

The potential effect on the resulting tuning curves from evaluating single-octave or multi-octave intervals was explored with the GAM approach similarly as above. The tuning values representing each note were calculated separately with each multi-octave condition with the algorithms described in Sec. III C. Note-dependent inverse variance weighting was also calculated separately for each multi-octave condition. The resulting tuning curves were then subjected to GAM analysis with the octave multiple as dependent variable. Due to the sparsity of the notes anchoring the tuning curves above five-octave intervals, the GAM model became unstable, and thus the inspection was limited between single and five-octave intervals. The comparison of resulting tuning curves with single- and multi-octave intervals is presented in Fig. 6.

The results of this analysis suggest that the span of evaluated octaves has a statistically significant effect on the shape of the resulting tuning curve averaged over all instruments ($F_{4,41} = 16.82$, $p < 0.001$). However, in this context it

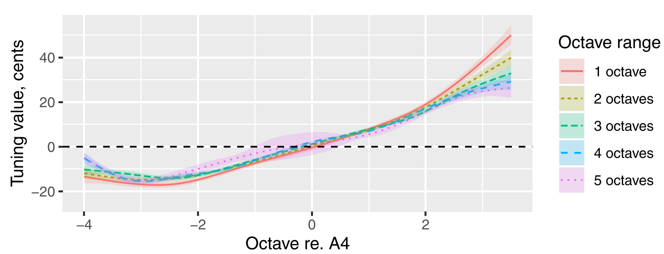

FIG. 6. (Color online) Tuning curves estimated with GAM for different evaluated octave spans.

is worth noting that the applied method of adjustment in the experiment may have an unintended effect. That is, the adjustment range of ±72 cents presented to the subjects with single-octave intervals was also in effect with multi-octave intervals. Particularly with wider multi-octave intervals, the available range imposed a restriction to certain evaluations of subjective interval enlargement. In contrast, the accumulation of single-octave stretching was less prone to this limitation, as the adjustment range caused some saturation only in the pitch extrema (see saturation effect in Fig. 2 at A0, A6, and E♭7). Consequently, this restriction caused some influence on the resulting tuning curves after the formulated tuning curve construction.

Since the restricting effect of the available tuning adjustment range was less substantial in the note range around the tuning centroid, the above analysis was repeated for tuning curves within octaves 2–6, thus, omitting the pitch extrema. These results again suggest that the tuning curves in the respective region differ with statistical significance depending on the octave multiples ($F_{3,13} = 3.9$, $p < 0.01$). However, it is worth noting that all the resulting tuning middle-region curves are nearly linear, and the effect of octave span on the average stretching slope of 7.82 cents/octave is at most 1.0 cent/octave.

E. Effect of listener background, age, and hearing

The participants in the listening experiment included a balanced selection of both professional and amateur musicians. GAMs were constructed for each subject group for comparison between possible differences in the average tuning curves due to the listeners' musical backgrounds.

The estimated tuning curves with the respective confidence intervals with similar inverse variance weighting as earlier are shown in Fig. 7. Visual inspection suggests that with professionals the pitch region between A2 and A6 is more linear and flatter than with amateurs. Conversely, the amateur curve shows more consistent overall shape above A2 while exhibiting a more pronounced J shape over the lowest octaves. However, the results from statistical analysis do not indicate a statistically significant difference between the shape of the curves ($F_{1,13} = 1.548$, $p = 0.21$). Therefore, it is conceivable that the musical background does not have a major influence on the shape of the tuning curve. Furthermore, the time to complete the entire experiment, including breaks, did not differ significantly between amateur (mean 3.0 h; std 0.7 h) and professional musicians (mean 3.11 h; std 1.5 h), according to Welch's t-test for unequal variances ($t_{34} = -0.22$, $p = 0.83$).

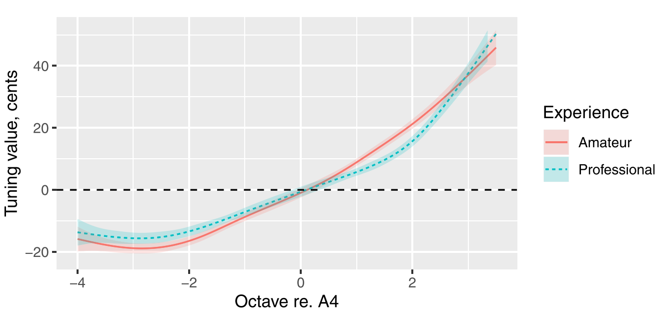

FIG. 7. (Color online) Comparison of average tuning curves estimated separately by data from professional musicians or amateurs.

The subject background data were also analyzed against the age groups between the median age (20–48 or 49+ years). The results indicate that the more senior subjects produced a smoother, more monotonous tuning curve, whereas the more junior listener group exhibited a pronounced J-shaped curve in the low-pitch region. A statistical test confirmed that the age groups yielded significantly different tuning curves ($F_{1,13} = 23.9$, $p < 0.001$).

Another investigation was conducted with the hearing screening data as the independent variable. One-third of the subjects had an elevated hearing threshold of at least 40 dB in one frequency band in either ear. The grand average curve calculated over these 12 subjects was compared with the respective curve produced by the other subjects. Unlike with age groups, the presence of hearing loss did not show statistically significant differences between the resulting tuning curves ($F_{1,13} = 1.05$, $p = 0.3$). It should be noted that the Pearson's correlation coefficient of 0.47 between the age of 49+ years and 40+ dB hearing loss was relatively high. Hence, these two comparisons are not entirely independent.

F. Instrument-dynamics interaction in octave stretching

The behavior of the octave enlargement effect as the interaction of instrument type and dynamic level is particularly interesting regarding practical conditions. However, according to our knowledge, GAM does not offer interaction models. Therefore, this aspect was explored with more conventional linear models. Here, we chose a subset of fundamental frequencies from either low or high registers and fit a linear approximation for the specific combination of conditions. Considering the moderate curvature within low and high octaves as seen in Fig. 5, a polynomial fit would have been more accurate, but would render the interpretation of results far more complex.

The null model was defined as tuning ∼ octave. The second model considering the effect of different instruments was defined correspondingly as tuning ∼ octave + octave:instrument, and the third model including the interaction effect by dynamic level was tuning ∼ octave + octave:instrument + octave:dynamics + octave:instrument:dynamics.

The high-pitch region was limited between notes A4 and A6, since only flute and strings could reach above those notes. The null model yielded a regression coefficient of 17.68 (cents per octave) with statistical significance ($F_{1,2158} = 199.19$, $p < 0.001$). The second model revealed that the effect of different instruments on the regression slopes did not reach significance ($F_{4,2154} = 1.67$, $p = 0.16$). However, the third model yielded a statistically significant but marginal effect by the dynamic level at the 0.05 level ($F_{2,2144} = 3.58$, $p = 0.029$). Fortissimo notes produced 0.57 cents/octave higher stretching of the tuning curve on average compared with pianissimo notes. The statistical significance between null and alternative models was evaluated with the

Chi-squared test. Compared with the null model, the second ($df = 4$, $p = 0.15$) and third models ($df = 10$, $p = 0.097$) did not explain the overall data substantially better.

Since the tuning curves for the flute and strings could be defined up to A7, a separate analysis with those instruments was conducted for the octave region A5–A7 in a similar fashion as above. With this combination, the effect of the instrument on the slope was found significant with the third model ($F_{1,1294} = 8.20$, $p = 0.004$) with the flute exhibiting 1.36 cents/octave steeper slope on the high-pitch region tuning curve in comparison to strings. Dynamics did not have a significant effect or interaction.

The low-pitch region of the tuning curves between E♭1 and A3 were subjected to the corresponding analysis. This excluded the flute from the analysis. A0 was omitted from the low-pitch region curves here for the identical dataset and also for clarinets, which lacked the lowest note in the original data. The instruments did not exhibit a statistically significant difference in the overall slope in the second model. In contrast, the third model showed a statistically significant difference compared with the null model with the Chi-squared test ($df = 8$, $p < 0.001$), and suggested that the interaction between instrument and dynamics has a significant effect on the tuning curve slope ($F_{6,2590} = 4.37$, $p < 0.001$). However, this effect was strongest in mezzoforte dynamics, with brass and double reeds increasing the slope noticeably over other conditions in dynamics extrema.

To summarize, a linear approximation model showed a marginal interaction effect between the tuning curves with instrument type and dynamic level in the high-pitch region. An interaction effect between instrument type and dynamics on the tuning curve slope was observed for the low-pitch region, although the effect was not linear with the increasing dynamic level.

V. DISCUSSION

Although the resemblance between our results and earlier findings is pronounced, some differences are apparent. In the low register below A2, the octave enlargement diminished, and in the case of clarinet even changed to an octave contraction (Fig. 5). As seen in Fig. 3, the stretching curve under A2 is quite stable and horizontal, whereas in earlier studies and in Railsback's curve, stretching in the low-pitch region is rather similar to that in the high-pitch region. Overall, in the lowest and highest pitch regions of the present study, deviation from the mathematical octave enlarged when the tones approached the limits of a human listener's ability to estimate pitches (A0 and E♭8).

In addition, the present results on non-significance by dynamic level on the tuning curve partially contrasts with the findings of Sundberg and Lindqvist,[7] who observed that the intensity of the stimuli influenced the SO. However, potential differences explaining this difference may lie in the naturally occurring dynamic spectra in the present stimuli and a narrower range of SPLs without uncomfortable extrema.

A comparison of the curve calculated from our results with the standard stretching curve of a Steinway D grand piano (New York; according to Tunelab software, Real-Time Specialties, Inc., MN) and Railsback's curve for the piano revealed that the shapes are different to some extent (Fig. 8). In both piano curves, the middle-pitch region is flattened around the tuning reference, and the curve is more S-shaped than the curve calculated from the present results. A possible explanation for this may have its roots in classic orchestration rules presumably affecting piano tuning. Typically, small harmonic intervals (seconds, thirds, even fourths) have been traditionally advised to avoid in chords in the low-pitch region due to sensory roughness (harmonics of the adjacent tones within the same critical band) and dissonance (no common lower harmonics). In contrast, in the middle-pitch region, a complex texture is generally a highly adopted manner to write music. In the high-pitch region, in turn, melodic lines and octave doublings are typical, especially in classic-romantic piano and orchestra repertoire.[36] In addition, Terhardt and Zick[10] verified this hypothesis in an experiment where they compared contracted, equal tempered, and stretched intonation in different musical contexts. According to their results, equal tempered or even contracted intonation was preferable in chords with medium or high spectral complexity. Hence, as such texture is commonly used in the middle-pitch region, this may explain, at least in part, the flattening of the stretching curve around the tuning reference. If the musical texture had a wider ambitus (i.e., large distance between simultaneously played notes), stretched intonation was rated best. That is, it is possible that this hypothesis should also be considered in our model even though it is not visible in our curves. Our experimental design was not suitable for evaluating this aspect due to lack of musical context. As mentioned before, musical background itself did not elicit statistically significant differences between groups. However, the most divergent individual results were measured among professionals.

Due to the present experimental design, where the lower tone was the reference tone and the higher tone was to be adjusted, some of the professional musician participants reported difficulty in recognizing the notes in the low register due to the equal-tempered scale used in the reference tones. The pitches of the low reference tones were considered too high compared with the intonation these

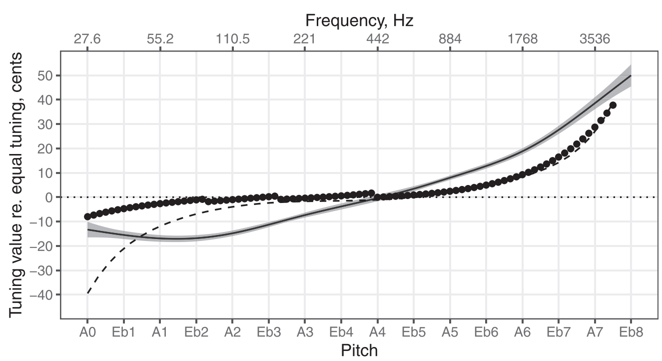

FIG. 8. Grand average of the current tuning curve (solid line) with corresponding confidence intervals (shaded area), and piano tuning curve for each tone in a Steinway model D grand piano (black bold dots) from Tunelab software (Real-Time Specialties, Inc., MN). The dashed line represents the Railsback curve adopted from Schuck and Young. The x axis (mathematical ET scale) is marked by a horizontal line with small dots (Ref. 16).

professional participants were used to in their daily work. Simultaneous stretched octave intervals played by instruments with harmonic spectra often cause beats in isolation or controlled scientific experiments. However, according to professional observations of the author J.J., it is rarely a problem in the real orchestra environments due to random phases and unstable intonation. Wide chords are particularly insensitive to beats. Beats usually occur when two instruments try to play in unison in the high register.

Differences between spectra may explain the J-shaped curve of clarinets. In the low register, the clarinet has a divergent spectrum from other instruments, where every second partial is attenuated. This may affect perceived pitch and will be investigated in our forthcoming study. To the best of our knowledge, no relevant studies about this phenomenon are available. It is also worth noting that intensive practicing of a musical instrument may influence the preferred tuning of scale or intervals, as in the case of the brass instruments, where just intonation is an essential part of the tuning system. In further studies, it would be relevant to broaden the level of musical background to non-musicians. In addition, the influence of Western music could be considered, and participants could also be from other musical cultures where the equal-tempered chromatic scale is not a de facto tuning standard.

VI. CONCLUSIONS

The present study explored the stretching of subjectively tuned octave and multi-octave intervals and the resulting tuning curves with complex tones of orchestra instrument spectra. Statistical analyses revealed that most instrument groups produce corresponding tuning curves, while the clarinet tones exhibit a slightly differing tuning curve in the low-pitch region. In general, the concept of a tuning curve appears well defined, as the corresponding average curves were attained through subjective evaluation multiple octave intervals.

The benefits for musicians, music educators, and also instrument makers would be remarkable if the stretched scale was accepted as an essential part of tuning philosophy. As discussed before, all experienced musicians tend to play upon principles of stretched tuning, although most of them unconsciously. If this approach is assimilated already in the early stages of music education, it may help to play better in tune and avoid tuning conflicts with the piano and other similar stretched tuned instruments. Some persistent tuning problems of certain woodwind instruments could also be better understood or corrected in practice with application of stretched tuning.

Apart from specialized tuning machines for piano tuning, chromatic tuners may have some temperaments available, but with cyclic octaves, they completely disregard the psychoacoustic phenomenon of octave enlargement. However, the present time, some tuning machines and applications have built-in stretching tuning curves (StroboPLUS HD and iStroboSoft, Peterson Electro-Musical Products Inc., IL), and are either fully programmable or have several presets available. Hopefully, true stretched tunings will find their way as a de facto feature in tuning equipment in the future.

ACKNOWLEDGMENTS

This research was supported by the Academy of Finland (Project No. 289300). The authors express their thanks to Dr. Jari Lipsanen at University of Helsinki for valuable advice on statistical methods.

[1]A. Bachem, "Tone height and tone chroma as two different pitch qualities," Acta Psycholog. **7**, 80–88 (1950).
[2]A. Bachem, "Chroma fixation at the ends of the musical frequency scale," J. Acoust. Soc. Am. **20**, 704–705 (1948).
[3]C. Stumpf and M. Meyer, "Maassbestimmungen über die Reinheit consonanter Intervalle (Measures concerning the purity of consonant intervals)," Beiträge zur Akustik und Musikwissenschaft **2**, 84–167 (1898).
[4]W. Ward, "Subjective musical pitch," J. Acoust. Soc. Am. **26**(3), 369–380 (1954).
[5]K. Walliser, "Über die Spreizung von empfundenen Intervallen gegenüber matematisch harmonischen Intervallen bei Sinustönen (On the spread of perceived intervals with respect to mathematic harmonic intervals in sinusoidal tones)," Frequenz **23**, 139–143 (1969).
[6]E. Terhardt, "Die Tonhöhe Harmonischer Klange und das Oktavintervall" ("The pitch of harmonic tones and the octave interval"), Acustica **24**, 126–136 (1971).
[7]J. E. Sundberg and J. Lindqvist, "Musical octaves and pitch," J. Acoust. Soc. Am. **54**(0001-4966), 922–929 (1973).
[8]W. Hartmann, "On the origin of the enlarged melodic octave," J. Acoust. Soc. Am. **93**(6), 3400–3409 (1993).
[9]P. A. Dobbins and L. L. Cuddy, "Octave discrimination: An experimental confirmation of the 'stretched' subjective octave," J. Acoust. Soc. Am. **72**(2), 411–415 (1982).
[10]E. Terhardt and M. Zick, "Evaluation of the tempered tone scale in normal stretched and contracted intonation," Acustica **32**, 268–274 (1975).
[11]G. Van Den Brink, "Octave and fifth settings for pure tones and residue sounds," in *Psychophysics and Physiology of Hearing*, edited by E. Evans and J. Wilson (Academic, London, 1977), pp. 373–379.
[12]L. Demany and C. Semal, "Harmonic and melodic octave templates," J. Acoust. Soc. Am. **88**(5), 2126–2135 (1990).
[13]B. S. Rosner, "Stretching and compression in the perception of musical intervals," Music Percept. **17**(1), 101–114 (1999).
[14]C. von Maltzew, "Das Erkennen sukzessiv gegebener musikalischer Intervalle in den äusseren Tonregionen" ("The recognition of successively presented musical intervals in the outer regions of the musical scale"), Z. Psychol. **64**, 161–257 (1913).
[15]O. Railsback, "Scale temperament as applied to piano tuning," J. Acoust. Soc. Am. **9**(3), 274–274 (1938).
[16]O. Schuck and R. Young, "Observations on the vibrations of piano strings," J. Acoust. Soc. Am. **15**(1), 1–11 (1943).
[17]O. Jorgensen, *Tuning: Containing the Perfection of Eighteenth-Century Temperament, the Lost Art of Nineteenth-Century Temperament and the Science of Equal Temperament* (Michigan State University Press, East Lansing, MI, 1991), pp. 739–745.
[18]J. Sundberg, *The Science of Musical Sounds (Cognition and Perception)* (Academic, San Diego, 1991), p. 274.
[19]J. F. Corso, "Scale Position and Performed Melodic Octaves," J. Psychol. **37**(2), 297–305 (1954).
[20]F. Fransson, J. Sundberg, and P. Tjernlund, "Statistical computer measurements of the tone-scale in played music," STL-QPSR **2-3**, 41–45 (1970).
[21]A. Rakowski, "Intonation variants of musical intervals in isolation and in musical contexts," Psychol. Music **18**, 60–72 (1990).
[22]E. Terhardt, "Stretch of the musical tone scale," available at https://www.ei.tum.de/fileadmin/tueifei/mmk/Personen/Terhardt/ter/top/scalestretch.html (Last viewed 20 October 2019).
[23]J. Schnupp, I. Nelken, and A. King, *Auditory Neuroscience: Making Sense of Sound* (The MIT Press, Cambridge, MA, 2011), p. 347.
[24]G. Langner and C. Benson, *The Neural Code of Pitch and Harmony* (Cambridge University Press, Cornwall, UK, 2015).
[25]A. J. Oxenham, "Pitch perception," J. Neurosci. **32**(39), 13335–13338 (2012).
[26]E. Terhardt, "Oktavspreizung und Tonhöhenverschiebung bei Sinustönen" ("Octave enlargement and pitch shift on sinusoidal tones"), Acustica **22**, 345–351 (1970).

[27]K. Ohgushi, "The origin of tonality and a possible explanation of the octave enlargement phenomenon," J. Acoust. Soc. Am. **73**(5), 1694–1700 (1983).
[28]S. Makeig and G. Balzano, "Octave tuning. Two modes of perception," in *Fifth Annual Symposium on the Psychology and Acoustics of Music* (1982).
[29]M. F. McKinney and B. Delgutte, "A possible neurophysiological basis of the octave enlargement effect," J. Acoust. Soc. Am. **106**(5), 2679–2692 (1999).
[30]A. Bell and W. W. Jedrzejczak, "The 1.06 frequency ratio in the cochlea: Evidence and outlook for a natural musical semitone," PeerJ **5**, e4192 (2017).
[31]A. Appunn, "Appun reed tonometer," in *The Collection of Historical Scientific Instruments of Harvard University*, available at http://waywiser.fas.harvard.edu/objects/3435/appun-reed-tonometer;jsessionid=4AA975134B38AEC767A9ED2658A27ABD (Last viewed 20 October 2019).
[32]E. M. Cramer and W. H. Huggins, "Creation of pitch through binaural interaction," J. Acoust. Soc. Am. **30**(5), 413–417 (1958).
[33]S. N. Wood, *Generalized Additive Models: An Introduction with R* (Chapman and Hall/CRC, Boca Raton, FL, 2006).
[34]C. Semal and L. Demany, "The upper limit of 'musical' pitch," Music Percept. **8**(2), 165–176 (1990).
[35]D. Pressnitzer, R. D. Patterson, and K. Krumbholz, "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**(5), 2074–2084 (2001).
[36]S. Adler, *The Study of Orchestration* (Norton, New York, 2002), p. 864.