Running title: Timbre blending during musical performance

Blending between bassoon and horn players: an analysis of timbral adjustments during musical performance

Sven-Amin Lembke, Scott Levine, and Stephen McAdams
Centre for Interdisciplinary Research in Music Media and Technology
Schulich School of Music, McGill University
Montreal, QC, Canada

29/05/2017

Abstract

Achieving a blended timbre between two instruments is a common aim of orchestration. It relates to the auditory fusion of simultaneous sounds and can be linked to several acoustic factors (e.g., temporal synchrony, harmonicity, spectral relationships). Previous research has left unanswered if and how musicians control these factors during performance to achieve blend. For instance, timbral adjustments could be oriented towards the leading performer. In order to study such adjustments, pairs of one bassoon and one horn player participated in a performance experiment, which involved several musical and acoustical factors. Performances were evaluated through acoustic measures and behavioral ratings, investigating differences across performer roles as leaders or followers, unison or non-unison intervals, and earlier or later segments of performances. In addition, the acoustical influence of performance room and communication impairment were also investigated. Role assignments affected spectral adjustments in that musicians acting as followers adjusted toward a 'darker' timbre, i.e., realized by reducing the frequencies of the main formant or spectral centroid. Notably, these adjustments occurred together with slight reductions in sound level, although this was more apparent for horn than bassoon players. Furthermore, coordination seemed more critical in unison performances and also improved over the course of a performance. These findings compare to similar dependencies found concerning how performers coordinate their timing and suggest that performer roles also determine the nature of adjustments necessary to achieve the common aim of a blended timbre.

Keywords

ensemble coordination, music performance, timbre, blend, spectrum

Among the many aims of orchestration, the combination of instruments into a blended timbre is one that is most relevant perceptually. Although decisions concerning orchestration can be primarily guided by personal preference, blend relies on a set of perceptual factors. It is commonly assumed to concern the auditory fusion of concurrent sounds into a single timbre, with the individual sounds losing their distinctness. Furthermore, it is thought to span a perceptual continuum from complete blend to distinct perception of individual timbres (Sandell, 1991; Kendall & Carterette, 1993; Sandell, 1995; Reuter, 1996; Tardieu & McAdams, 2012; Lembke & McAdams, 2015). Perceptual cues that are favorable to blend range from synchronous note onsets and pitch relationships emphasizing the harmonic series, to instrument-specific acoustical traits. Concerning pitch relationships, higher blend is achieved for unison than for non-unison intervals (Kendall & Carterette, 1993). Whereas dissonant pitch intervals exhibit greater frequency divergence between harmonics that may render the identities of constituent instruments in a mixture more distinct, combinations in highly consonant intervals (octaves, fifths) can be assumed to be more blended. For the latter, auditory fusion can be further enhanced by parallel movement of voices (Bregman, 1990). For all non-unison intervals, certain combinations of instruments can be expected to lead to higher degrees of blend than others, which may influence the instrumentation choices orchestrators make.

With respect to acoustic traits, previous studies have shown spectral properties to have the strongest effect on blend between sounds from sustained instruments. The global spectral shape of many wind instruments has been shown to be largely invariant with respect to pitch and may also bear prominent features such as spectral maxima (Lembke & McAdams, 2015). These maxima are also termed *formants*, in direct analogy to the pitch-independent spectral maxima found in human voice production (Fant, 1960). Previous explanations that relate blend to spectral features are either based on global spectral characterization or focus on local, prominent spectral traits. The global and more general hypothesis was established from studies for instrument dyads, in which the spectral centroids of individual instruments were evaluated. The spectral centroid represents the global, amplitude-weighted frequency average of a spectrum. It has been shown that higher degrees of blend are obtained when the sum of the spectral centroids of the constituent instruments are lower (Sandell, 1995; Tardieu & McAdams, 2012). The alternative hypothesis argues that localized spectral features influence blend, more specifically, concerning formant relationships between instruments: when two instruments exhibit coincident formant locations, high blend is achieved, whereas increasingly divergent formant locations decrease blend, as the individual identities of instruments are thought to become more distinct (Reuter, 1996).

Lembke & McAdams (2015) followed up on the formant hypothesis by studying frequency relationships between the most prominent *main* formants. The investigation considered dyads of recorded and synthesized instrument

sounds. The recorded sound remained a static reference and the synthesized sound was varied parametrically with respect to its formant frequency. For the instruments with prominent formant structure, namely bassoon, (French) horn, trumpet, and oboe, blend was found to decrease markedly when the synthesized main formant exceeded that of the reference, whereas comparably high degrees of blend were achieved if the synthesized formant remained at or below the reference. This rule proved to be robust across different pitches, with the exception of the highest instrument registers, and even applied to non-unison pitch intervals. However, this rule relies on one instrument serving as a reference, which raises the conundrum of which of two instruments in an arbitrary combination would function as the reference. The answer may lie in musical practice: either the instrument leading the joint performance or the one with a more dominant timbre could assume this function.

In musical practice, achieving blended timbres involves two stages: its conception and its realization. Blend is first conceived by composers and orchestrators, who lay out the foundations by providing necessary perceptual cues, i.e., ensuring that musical parts have synchronous note onsets and pitch relationships favorable to blend, with the parts being assigned to suitable instrument combinations. The successful realization of blend as perceived by listeners still depends on musical performance, which necessitates precise execution by several performers with respect to intonation, timing, and likely also coordination of timbre. Previous research precluded the influence of performance by relying on stimuli that were mixed from instrument sounds that had been recorded in isolation, with there being a single exception (Kendall & Carterette, 1993) in which dyad stimuli had been recorded in a joint performance (Kendall & Carterette, 1991). The interaction between performers may in fact influence blend in a way that previous research has not considered. For instance, differences between performer roles could provide answers to the question of a certain instrument serving as a reference.

Musical performance

Psychological research on musical performance has primarily investigated temporal properties. Although past investigations have focused on note synchronization and timing between performers (Rasch, 1988; Goebl & Palmer, 2009; Keller & Appel, 2010) as well as related motion cues (Goebl & Palmer, 2009; Keller & Appel, 2010; D'Ausilio et al., 2012), performer coordination with respect to timbral properties remains largely unexplored (Keller, 2014; Papiotis et al., 2014). Rasch (1988) established that a certain degree of asynchrony between performers is common and practically unavoidable, whereas perceptual simultaneity between musical notes is still conveyed. For example, typical asynchronies between wind instruments (e.g., single and double reed) performing in non-unison are reported as falling within 30-40 ms. Moreover, the asynchronies relate to different roles assumed

by musical voices, e.g., the melody generally precedes bass and middle voices.

Two studies investigated the relationship between two pianists being assigned performer roles as either *leader* or *follower*. In one study, followers exhibited delayed note onsets relative to leaders (Keller & Appel, 2010), whereas in the other, followers displayed a higher temporal variability, thought to be linked to a strategy of error correction relative to leaders (Goebl & Palmer, 2009). In addition, the second study showed that under impaired acoustical feedback, performers increasingly relied on visual cues to maintain synchrony. Investigations with a sole focus on performance-related factors within the auditory domain would therefore need to prevent visual communication between musicians.

Role dependencies between performers are indeed common to performance practice. They have been investigated for larger ensembles (D'Ausilio et al., 2012) and have been discussed in terms of *joint action* (Keller, 2008), in which they may modulate how performers rely on cognitive functions such as anticipatory imagery, integrative attention, and adaptive coordination. In terms of musical interpretation, leaders commonly assume charge of phrasing, articulation, intonation, and timing, whereas followers "adapt their own expressive intentions to accommodate or blend with another part" (Goodman, 2002, p. 158). It therefore appears plausible that the performance of blended timbre may similarly rely on role assignments between musicians. For instance, when two instruments are doubled in unison, one of them assumes the leadership in performance, toward which followers may orient their timbral and timing adjustments. In addition, these adjustments may continually be refined, as it likely takes some time for both musicians to improve their coordination, given their individual roles and respective performance goals.

The current study explores what timbral adjustments are employed in achieving blend and how these interact in a performance scenario with two musicians. A set of acoustic measures monitors the spectral change and potential covariates that are assumed to be related to timbral adjustments. In addition, performances are also evaluated through musicians' self-assessments. Besides timbral adjustments, performances naturally also involve aspects related to timing, intonation, and adjustment of dynamics. Intonation has not been previously discussed as relating to blend, likely due to past research having precluded performance-related aspects, but reports from performers argue that correct intonation aids blending. Given the emphasis on timbre, however, performer coordination with respect to synchronization and intonation remains outside the focus of the current study. Moreover, they represent aspects that are important to accurate delivery of musical performance in general, which greatly limits the extent to which they can be varied independently to affect blend. As a result, the emphasis in this article lies on the spectrum, which likely governs instrumentation choices composers

make and relates to the timbral adjustments over which performers have independent control.

The investigation considers a realistic account of factors encountered in musical practice and situates musicians in an approximation to the ecologically valid setting of a concert hall, realized through controlled and reproducible virtual performance environments. In concert halls, the *coloration* of instrument timbre as a function of relative position inside the room has been reported to be perceptible (Goad & Keefe, 1992), which would similarly extend to differences between rooms. Furthermore, an impairment of the acoustical communication between musicians (Goebl & Palmer, 2009) may be relevant to the performance of blended timbre as well. Because the investigation considers a potential effect of performer roles, an instrument combination should be chosen that allows for sufficient timbral coordination, i.e., by avoiding situations in which one instrument's timbre dominates the other when a change in role assignments is unlikely to overcome the strong timbral mismatch. An instrument combination that is widely used in the orchestral repertoire is bassoon and horn. Orchestration treatises discuss these two instruments as forming a common blended pairing (Rimsky-Korsakov, 1964; Koechlin, 1954), with these observations reflected in findings of high degrees of blend in perceptual investigations (Sandell, 1995; Reuter, 1996). The horn is often considered an unofficial member of the woodwind section, bearing a timbral versatility that succeeds in blending with woodwinds, brasses, and even strings, which suggests that, at the very least, it should succeed in bridging timbral differences with the bassoon.

In summary, this investigation tests several hypotheses based on the following experimental variables or factors (set in *italics* and capitalized). It is expected that musicians will perform differently as leaders than as followers, with those in the *Role* of followers adjusting their timbre to that of the leader. Unison *Intervals* are hypothesized to yield higher perceived blend than the non-unison case, as well as possibly showing more coordination between instrumentalists. Furthermore, the coordination between performers is predicted to increase throughout a performance, i.e., it should be higher in a later than an earlier musical *Phrase*. With respect to the influence of acoustics, differences between *Rooms* may affect the degree of coordination between performers to some extent, although it is not clear in what way. Finally, given an assumed stronger dependency of followers on leaders than vice versa, performances in which leaders lack acoustical feedback from followers are not expected to differ from the case with unimpaired *Communication*.

Acoustic measures for timbre adjustments

Our acoustical analysis of instruments focuses on the spectral envelope, which represents the envelope or profile outlined by the partial tones contained in an instrument's spectrum (Rodet & Schwarz, 2007). Unlike conventional Fourier spectra, which characterize spectral fine structure by

delineating individual partial tones and the gaps between them, a spectral envelope is a smooth, continuous function approximating the broader spectral structure of instruments, e.g., revealing the presence of formants, which one might conceive of as the resonant structure that shapes the amplitudes across frequencies. Spectral envelopes can be determined for audio signals across their time course (Villavicencio et al., 2006) or they can concern pitch-generalized descriptions from a compilation of spectra obtained across entire pitch ranges of instruments (Lembke & McAdams, 2015). With regard to the latter, bassoon and horn bear a high resemblance, as illustrated in Figure 1 for the dynamic marking *piano*. As their most prominent traits, main formants are located around 500 Hz and can be characterized by the frequency F_{max} (solid red line) corresponding to the maximum magnitude and the frequency above F_{max} where the magnitude has decreased by 3 dB, termed the upper frequency bound F_{3dB} (dashed red line). Both instruments' main formants exhibit similarities, with their F_{max} differing by only about 80 Hz, whereas their F_{3dB} lie much closer. In addition, the spectral centroids *SC* (solid blue line) are located in the vicinity of the main formants, showing the global spectral distribution to be strongly influenced by the prominence of the main formants. Still, the horn exhibits a slightly broader, more dominant main-formant region, which may equate to a similar difference in timbral dominance.

[Insert Figure 1 about here]

Although the pitch-generalized description in Figure 1 approximates the instruments' structural invariants, i.e., related to what informs orchestrators in their choice of instruments, in practice these structural constraints still allow for a certain degree of timbral variation that musicians can exploit. Because wind instruments act as acoustic systems in which all sound originates from common structural elements (e.g., mouthpiece, resonator tube), timbral adjustments are expected to be inherently linked to the primary parameters of sound excitation performers focus on, namely, pitch and dynamic intensity. For both instruments, blend-related adjustments of timbre can be assumed to relate to spectral changes, which can be monitored by evaluating time-variant spectral envelopes (e.g., by way of *True Envelope* (TE) estimation, Villavicencio et al., 2006), again employing the descriptive measures F_{max}, F_{3dB}, and *SC*. An example is given in Figure 2 (see color plate section), showing a horn playing an ascending A-major scale over two octaves, visualized as a spectrogram of TE estimates across time frames. Apart from the spectral descriptors F_{max}, F_{3dB}, and *SC*, the figure includes the temporal evolution of pitch and dynamics, represented by the fundamental frequency f_0 (white curve) and the relative sound level L_{rms} (level sum across all frequencies: separate horizontal strip at the bottom), respectively. Gaps in the spectral descriptors F_{max} and F_{3dB} (red curves) are due to unreliable detection of formants.

[Insert Figure 2 about here. Note: In color!]

From a preliminary qualitative investigation with bassoon and horn players, the timbre variability at the players' control was found to be greater for horn than for bassoon. For the latter, the location and shape of the main formant is relatively fixed, with spectral changes primarily affecting the magnitudes of higher frequency regions relative to the main formant, whereas the structural constraints of the horn allow for greater changes to main-formant location and shape, as also becomes apparent in Figure 2. Musicians reported that during performance, the greatest timbre change could be achieved by varying dynamics, which suggests a dependency between them. The identification of perceived dynamic markings has been shown to be mediated by both timbre and sound level (Fabiani & Friberg, 2011), which argues that when performers adjust dynamics, both timbre and the sound level (L_{rms}) are affected.

Apart from dynamics, pitch presents another source of covariation with spectral measures, with pitch being expressed through the fundamental frequency (f_0) for harmonic sounds. In Figure 2, all spectral measures show some variation as pitch ascends, which can be quantified descriptively by the linear correlation coefficient (Pearson's *r*): The strongest covariation with f_0 is apparent for *SC*, $r = .92$, whereas the correlation with main-formant measures is less pronounced, $r < .40$, with F_{max} and F_{3dB} meandering around idealized average values. Given these differences in covariation with f_0, the two types of spectral measures seem to capture independent contributions of timbral change. It is important to note that even f_0 and L_{rms} yield a clear degree of correlation, $r = .72$, with about 10 dB of level change across the two octaves. In orchestration practice, this correlation corresponds to the notion of *pitch-driven dynamics*, with experimental evidence showing that ascending pitch contour can enhance the identification of changes in dynamics, e.g., *crescendo* (Nakamura, 1987). In summary, this preliminary investigation suggests that timbral adjustments should be evaluated by way of combined measures of spectral variation and other potential factors of covariation, such as pitch and dynamics.

Method

Participants

Sixteen musicians were recruited primarily from the Schulich School of Music at McGill University and the music faculty of the Université de Montréal. The bassoonists, three female and five male, had a median age of 21 years (range 18-31). The hornists, six female and two male, had a median age of 20 years (range 17-44). Across both instruments, 10 participants considered themselves professional musicians, and overall, the musicians reported playing or practicing their respective instruments for the median

duration of 21 hours per week (range 5-35). All musicians were paid for their participation and provided informed consent. The study was reviewed for ethical compliance by the McGill Research Ethics Board.

Stimuli

Three musical parts were investigated, all taken from a single excerpt in Mendelssohn-Bartholdy's *A Midsummer Night's Dream*, Op. 61, No. 7 (measures 1-16). The chosen instrument combination is featured prominently in this musical passage. In a thin orchestral texture, low strings, second horn, and clarinet establish the harmonic structure through long, separated notes, while two bassoons accompany a solo horn melodically. In the absence of other salient voices, the combination of bassoons with horn can therefore be thought to aim for a homogeneous, *blended* timbre. All parts were transposed by a fifth down to A major from the original key of E major, to reduce the impact of player fatigue through repeated performances in high instrument registers, at the same time ensuring little change in key signature. The transposed parts are shown in Figure 3. The melody, voice A, was used for unison performances, whereas voices B and C served as non-unison material. Across the different experimental conditions, each voice was played by both instruments, regardless of whether a voice had been assigned to only one particular instrument in the original score.

[Insert Figure 3 about here.]

Although the musicians played in separate rooms in order to record their individual sounds, they heard themselves and the other player over headphones in a simulated virtual-acoustics environment, which allowed the control over acoustical factors (see *Design*). The simulation was achieved through binaural reproduction (Paul, 2009) using real-time convolution of the instruments' source signals with individualized binaural *room impulse responses* (RIRs). Each musician's performance was captured through an omnidirectional microphone (DPA *4003-TL*). Both microphone signals were routed to a control room, where preamplification gain was digitally matched for both performers. The analog signals were converted to 96 kHz / 24-bit PCM digital data, recorded at full resolution for later acoustical analysis and at the same time fed into separate convolution engines that processed the source signals with customized RIRs, based on the manipulation of acoustical factors. Individualized binaural signals were then fed to headphones for each performer. Headphone amplifier volume was held constant, as were the circumaural closed-ear headphones (Beyerdynamic *DT770*). A latency inherent to the convolution delayed the arrival of the simulated room feedback by about 8.4 ms, affecting both performers equally. The RIRs had been previously collected in real concert venues and were measured with a binaural head-and-torso system (Brüel & Kjaer *Type 4100*), excited by a

loudspeaker (JBL *LSR6328P*) positioned to emulate the instruments' main sound-radiation directivity (Meyer, 2009).

In the simulated environment, musicians would hear themselves and the other musician in a common performance space, which provided realistic room-acoustical cues (e.g., room size, its reverberation characteristics, relative spatial positions of players). The instrument locations were based on a typical orchestral setup: horns on the conductor's left front side and bassoons on the conductor's right front. For instance, hornists heard themselves in direct proximity and the bassoonist towards their left, at a distance of 3.6 m, whereas the bassoonists' viewpoint was reversed in orientation. In order to take these individual viewpoints into account, i.e., as performers heard themselves (*self*) and the other musician (*other*), the acoustical analyses of performances considered the individualized binaural signals. Although four possible binaural signal paths resulted from a performer having two ears and hearing two sources at the *self* and *other* positions, only two paths were considered for simplicity: *self* considered the ear facing away from the other performer, and *other* considered the ear closer to the other performer.

Design

Performances were studied as a function of musical and acoustical factors using a repeated-measures design to rule out confounding individual differences for instruments and playing technique or style with the investigated effects.

Musical factors.

Three independent variables considered the performer role, the influence of different musical voice contexts, and performance differences across time. For the Role factor, one instrumentalist was assigned the role of *leader*, while the other performer acted as *follower*, i.e., took on an accompanying role. According to the Interval factor, musicians either performed a melodic phrase in *unison* (voice A in Figure 3) or a two-voice phrase in *non-unison* (B and C); in non-unison, the top voice (B) was assigned to the leader. The Phrase factor divided the musical excerpt into two, with the separation occurring right before beat three of measure eight (see the 'V' in Figure 3). This separation yielded two musical phrases of identical length consisting of similar musical material, more so for unison than for non-unison parts.

Acoustical factors.

Two other variables investigated effects for communication directivity between performers and the room-acoustical properties of performance venues. The Communication factor assessed the influence of whether both performers were able to hear each other or whether only the follower could hear the leader, denoted *two-way* or *one-way*, respectively. For the Room factor, the influence of acoustics was assessed for two different performance spaces: musicians were simulated as performing in either a large,

multipurpose performance space (RT_{60} = 2.1 s, time for reverberation to decrease by 60 dB) or in a mid-sized recital hall (RT_{60} = 1.3 s).[1]

Procedure

The experiment was conducted in two research laboratories at the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) at McGill University. Separate laboratory spaces were called for in order to create individual acoustical environments for each participant, ensuring the capture of separate source signals as well as preventing visual cues between performers. Each performance laboratory was treated to be relatively non-reverberant, with $RT_{60} < 0.5$ s. Performers received instructions and provided feedback through dedicated computer interfaces. Musical notation for all three parts was provided on a music stand, and performances were temporally coordinated by a silent video of a conductor. With both performers seated on chairs, the stand was positioned to allow the performer's field of view to cover both the musical notation and the conductor, arranged similarly to the binaurally simulated orchestra situation, i.e., the stand slightly to the right of the conductor as seen from a hornist and to the left for a bassoonist. The video was recorded in advance by having an experienced conductor (with baton) outline the metrical structure of the musical excerpt, including gestures related to phrasing and articulation. He used a constant reference tempo of 58 beats per minute.

A pair of one bassoon and one horn player was tested in a single experimental session, being instructed to perform together to achieve the highest degree of blend possible. They performed three repetitions of 16 different experimental conditions (four factors by two treatment levels, excluding Phrase), leading to a total of 48 experimental trials. The experiment lasted around two hours in total, including a break scheduled after half of the trials. To avoid disorientation of musicians through strongly varying performer-role and voice assignments, the musical factors were grouped in separate blocks. Participants assumed the role of either leader or follower throughout the first or second half of the experiment. Furthermore, shorter eight-trial blocks grouped conditions based on voice assignment (e.g., four unison trials, another four non-unison), with the repetitions occurring after each block. For instance, a given participant would begin as leader for 24 trials, performing the first repetition of four unison trials, then proceed to four non-unison trials, followed by the second repetition of the same four unison trials, etc. The four possible block-ordering schemes were counterbalanced across all participants and instruments. The acoustical-factor combinations were nested in sub-blocks of four trials and randomly ordered. Three practice trials were conducted under the guidance of two experimenters

[1] The performance venues correspond to the *Music Multimedia Room* and *Tanna Schulich Hall*, respectively. Both are located at the Schulich School of Music, McGill University. More details under http://www.mcgill.ca/music/node/48232. (Last accessed on May 18, 2017.)

ahead of the main experiment, involving the experimental conditions from the first block of four trials.

A single experimental trial consisted of three stages: preparation, performance, and ratings. During preparation, musicians were asked to prepare the assigned musical parts and individual performer roles, while being able to hear themselves in the current simulated room environment. After both participants indicated being prepared, the actual performance commenced and once it ended, each participant judged their individual experience of the performance by providing two ratings. The first rating assessed how well they thought they had individually *performed* given their assigned role on a continuous scale with the verbal anchors *very badly* and *very well*. The second rating concerned the perceived degree of achieved *blend* with the other performer on a continuous scale with the verbal anchors *low blend* and *high blend*.

Acoustic measures

In addition to the behavioral ratings, several acoustic measures accounted for blend-related timbre features and were evaluated as time series. Timbral adjustments were evaluated through spectral descriptors and also monitored through the covariate measures pitch and dynamics. Two additional cues important to blend, namely, intonation and synchrony, were initially considered in order to allow their influence to be filtered out subsequently. Time series were analyzed with respect to the time-averaged magnitude of an acoustic measure, its temporal variability during performance, and its temporal coordination between performers. Therefore, each measure yielded three corresponding dependent variables (DVs).

All acoustic measures were based on spectral analyses across the time course of performances, for which short-time Fourier transforms (STFT) and further derived representations were computed using dedicated software (AudioSculpt/SuperVP, IRCAM, Paris). STFT was based on the fast Fourier transform (FFT), using Hann-windowed analysis frames consisting of 7620 samples, FFT length of 8192 bins, and an overlap of 25% between successive frames. Given the sampling rate of 96 kHz, this corresponded to a frequency and time resolution of 11.7 Hz and 19.8 ms, respectively. Pitch detection employed harmonic analysis of the STFT spectra (Doval & Rodet, 1991), with the identified fundamental frequency f_0 configured to fall within the possible range $f_0 \in$ [92.5, 370] Hz, which reflected the pitch range across all parts expanded by a whole tone on each end. The f_0 estimates provided by AudioSculpt were complemented by corresponding confidence scores, i.e., the likelihood for identified harmonics to be linked to f_0, which in turn were used to discard time frames falling below 80% confidence from further analysis for all measures. This elimination improved the reliability of both f_0 and spectral measures. Based on the remaining STFT frames, spectral envelopes were obtained through *True Envelope* (TE) estimation (Villavicencio et al., 2006). The TE algorithm applied iterative cepstral

smoothing on STFT-magnitude spectra, yielding individual spectral-envelope estimates per time frame, based on a constant cepstral order oriented at $f_0 \leq 300$ Hz. Then, a formant-analysis algorithm evaluated the spectral envelopes, identifying main formants (F_1), which were quantified in terms of frequencies characterizing their maximum F_{max} and the upper bound F_{3dB}, as well as computing the spectral centroid *SC* (Peeters et al., 2011). The spectral envelopes also served to quantify dynamics by determining relative, root-mean-square (RMS) power levels L_{rms}, which corresponded to the level summed across all frequencies of the spectrum.

As the raw time-series data for the measures exhibited some fine temporal variation and occasional outliers, some prior data treatment was needed. All measures were smoothed by a weighted moving-average filter. Weights were based on the f_0-confidence scores, assuming that higher confidence reflected a more robust and reliable parameter estimate. Smoothing used a sliding-window duration of 475 ms, which corresponded to an eighth note at the performed tempo. Especially for horn signals, the automated formant detection at times led to erroneous estimates, which could be identified and eliminated. Prior to smoothing, the main-formant descriptors F_{max} and F_{3dB} were filtered for outlying values that lay beyond an octave below and two-thirds of an octave above their time-averaged median value, because unrelated spectral features beyond these frequencies were occasionally classified as the main formant. Deemed an artifact of cepstral smoothing, the TE estimates for horns sometimes also exhibited spectral-envelope maxima at 0 Hz, in which case formant identification failed. Therefore, resulting gaps for F_{max} greater than two metrical beats were replaced by f_0 values, serving as the lowest tonal signal components. The corresponding F_{3dB} values were determined from the replaced F_{max}. The final step of data treatment ensured that the measures yielded values across all analysis frames of a performance, allowing comparisons between performers across all time points. This was achieved through linear interpolation of all remaining gaps to a reference time grid. Extrapolation was applied for values missing at the edges, which rarely exceeded a quarter-note duration (e.g., delayed entry of the first note or the final note not being held for its entire duration).

The investigation focused on timbral adjustments as reflected in spectral changes. However, not all spectral changes were necessarily related to the intent to achieve blend. Performer actions related to errors in intonation or timing could also have evoked a certain degree of spectral change. Therefore, the performances were filtered for cases in which bad intonation and/or synchrony were apparent. Intonation was measured by comparing f_0 between performers, expressed as the relative deviation in cents. For unison, this characterized deviations from a f_0 ratio of unity; for non-unison, the deviation considered f_0 ratios of the corresponding intervals in equal temperament. Asynchrony could also be assessed through the intonation measure, because asynchronous note entries also introduced substantial deviations from perfect intonation for the duration by which they were offset from synchrony. The

time series for all measures retained only values falling within the intonation range of ±25 cents, which corresponds to musically acceptable intonation (Rakowski, 1990). Unlike intonation and timing, pitch (f_0) and dynamics (L_{rms}) were intrinsically related to the spectral measures and could not be directly excluded from further analysis, but were instead monitored for similar trends along the spectral measures' time series. The influence of f_0 was twofold: First, systematic differences in f_0 between the musical parts were likely reflected in deviations between unison and non-unison performances. Second, f_0 also varied over time, and all spectral measures covaried with f_0 to some extent. By taking residuals (ε) from the linear regression of the f_0 time series onto the time series of each of the three spectral measures and adding the residual scores to the spectral time-series means, the linear covariation with f_0 over the parts could be removed. This procedure yielded the *residual* measures εF_{max}, εF_{3dB}, and εSC.

The performance analysis considered individual performers and evaluated each acoustic measure with three DVs. The first DV quantified the acoustic measure's average magnitude, using the *median* across time values. The second DV assessed the temporal variability along a measure, expressed as a robust *coefficient of variation* (CV): the ratio between interquartile range and median. The third DV assessed the temporal coordination between performers, evaluating the maximum *cross-correlation coefficient* (XC) for their time series.[2] Due to the expected covariation with f_0, the XCs for the spectral measures were assumed to be inflated by the inherent similarity in f_0 profiles between parts A and A (unison), and even B and C (non-unison). Therefore, this DV considered the residual measures (ε), whereas the remaining DVs were based on the original acoustic measures. Furthermore, in considering the individual viewpoints of performers within the binaural simulation, the DVs evaluating median and CV were based on time series for the binaural signal *self*, whereas the DV evaluating XC compared *self* with *other*.

Results

The presentation of results focuses on the hypotheses established in the introduction, which were tested by a total of five factors, namely, Role, Interval, Room, Communication, and Phrase, with two treatment levels each. In the experiment, performances across the 16 factorial combinations (excluding Phrase) were repeated three times. The subsequent analysis retained only the two 'best' repetitions per participant pair, i.e., those that

[2] Although cross-correlation time lags were also evaluated, no evidence for relative delays in coordination was found across all measures. For instance, L_{rms} displayed a median lag of 0 ms across all conditions and both instruments, with the interquartile range also being 0 ms, showing hardly any variation along this measure. *SC* exhibited a median lag of 0 ms with an extremely wide interquartile range of 871 ms, which reflects little agreement across participants.

yielded the highest self-assessed performance ratings, which needed to reflect agreement between the two participants performing together. Out of three repetitions, at least one found mutual agreement between both performers as to having been rated among the highest two. If there was no further mutual agreement, the repetition yielding the higher average rating across performers was taken. Some unforeseen technical issues during two experimental sessions rendered data for a total of five trials unusable. Fortunately, this affected only one repetition per experimental condition, allowing the remaining two repetitions to be used. In the analyses, separate performances were considered as independent cases, i.e., corresponding to a total of 16 cases (eight performers × two repetitions) per instrument.

Analyses of variance (ANOVAs) tested effects across the within-subjects musical and acoustical factors. The within-subject residuals yielded slight departures from a normal distribution (Shapiro-Wilk test). Based on the known robustness of ANOVA to violations of normality for equal sample sizes (Harwell et al., 1992), the use of ANOVA was considered justified for DVs exhibiting less than 10 violations over all 32 factor cells, which all reported statistical effects fulfilled. Furthermore, the two *Instrument* groups could be implemented as a between-subjects factor if both groups exhibited similar variances. This condition was fulfilled for the behavioral ratings, as both groups of players used identical rating scales and did not exhibit systematic differences in their ratings. The acoustic measures, however, exhibited clear violations (Levene's test), brought about by consistent differences in their acoustical characterization. As a result, the acoustic measures involved separate ANOVAs by instrument. In line with the use of ANOVA for repeated measures, reported main effects consider statistics for within-subjects differences between two levels of a single factor, i.e., means and standard errors across participants for individual differences along the factor in question. For a quantification of several DVs in terms of group means for individual factor cells, please refer to two tables in the supplementary materials **[Insert link here]**.

Behavioral ratings

Participants provided two ratings quantifying their perception of *blend* and assessment of their own *performance* given their assigned role. As the ratings applied to entire performances, mixed ANOVAs included the four within-subjects factors Role (leader, follower), Interval (unison, non-unison), Room (large, small), and Communication (two-way, one-way), with Instrument (bassoon, horn) forming a between-subjects factor.

For *blend* ratings, performers acting as leaders did not provide ratings for the impaired acoustical feedback as they were unable to hear the follower. To work around these missing values, separate ANOVAs evaluated two subsets of the blend ratings, which each excluded one of the problematic factors. The first only considered unimpaired feedback across the remaining within-subjects factors Role × Interval × Room; the second comprised only

performers acting as followers across Interval × Room × Communication. Both analyses suggested that performances were perceived as more blended in unison than in non-unison, without other factors interacting. Whereas performances under unimpaired communication yielded clear trends for higher blend in unison, $F(1,30) = 19.40$, $p < .01$, $\eta_p^2 = .39$, analysis of only followers' ratings resulted in only marginally higher blend ratings for unison, $F(1,30) = 3.94$, $p = .06$, $\eta_p^2 = .12$. In numerical terms, the observed blend-rating differences between unison and non-unison conditions amounted to a mean within-subject difference of about .04 (standard error .01) on a full scale range of [0, 1]. In summary, performances under unimpaired communication led to higher blend ratings for unison conditions, although the exclusion of leaders' ratings or the inclusion of ratings for impaired communication may have compromised this effect.

Performance ratings only led to a marginally significant main effect for Interval, $F(1,30) = 3.90$, $p = .06$, $\eta_p^2 = .12$, but this factor still yielded two-way interactions with Role, $F(1,30) = 6.43$, $p = .02$, $\eta_p^2 = .18$, and Communication, $F(1,30) = 4.70$, $p = .04$, $\eta_p^2 = .14$. Figure 4 presents differences between Roles (rating as leader minus rating as follower) and Communication direction (one-way minus two-way condition). As is apparent in Figure 4 (top panel), the first interaction involved musicians rating themselves as having performed their role better as followers than as leaders in unison conditions, with the inverse relationship holding for non-unison performances. The second interaction (Figure 4, bottom panel) suggested that in unison performances, musicians rated their performances higher for unimpaired, two-way communication, whereas the ratings for non-unison performances appeared to be unaffected by communication directivity.

[Insert Figure 4 about here]

Two additional interactions involved differences between instruments. Figure 5 presents the differences between Instruments (bassoon minus horn) and Roles (leader minus follower). As illustrated in Figure 5 (top panel), a two-way interaction with Role, $F(1,30) = 6.49$, $p = .02$, $\eta_p^2 = .18$, yielded higher performance ratings for bassoons than horns in the role of followers, whereas no difference between instruments was found for leaders. The same interaction suggested that bassoonists provided higher ratings as followers than as leaders (Figure 5, bottom panel), with the opposite applying to horns. A related three-way interaction (Figure 5, bottom panel) added the influence of the Room factor, $F(1,30) = 4.22$, $p = .05$, $\eta_p^2 = .12$. For bassoons, the difference between roles became larger in the smaller room, whereas for horns, the role difference appeared to be limited to just the smaller room.

Overall, these interdependencies suggest that communication impairment had a stronger effect on unison performances and that followers were more satisfied with their performances than were leaders. Differences between

instruments and across roles could be related to instrument-specific issues concerning playability of the corresponding parts. Furthermore, the less reverberant acoustics of the small room seemed to affect performances (or their evaluation) more critically.

[Insert Figure 5 about here]

Acoustic measures

The way in which bassoonists and hornists coordinated their playing to achieve blend was analyzed across the time course of performances by taking several acoustic measures into account. The analysis approach examined performer coordination as a function of the musical and acoustical factors being studied.

Figure 6 (see color plate section) visualizes a single performance by one bassoon and one horn player in two spectrograms obtained through TE estimation. The superimposed curves represent the time courses for all acoustic measures, F_{max}, F_{3dB}, SC, and f_0, and the separate horizontal strip at the bottom traces the temporal evolution of L_{rms}. In this example, the unison part was performed under normal, two-way communication in the larger room, with the bassoon acting as leader. This example also considers the bassoon's viewpoint, i.e., involving binaural signals for bassoon and horn as heard from the *self* and *other* positions, respectively. Three DVs were derived from each measure — median, CV, and XC — and were analyzed in repeated-measures ANOVAs investigating the factors Role, Interval, Room, Communication, and Phrase.

[Insert Figure 6 about here. Note: In color & landscape orientation!]

Because the acoustic measures and associated DVs were quantified along physical scales or quantities derived from them, statistical effects were also evaluated against psychoacoustically meaningful thresholds. For median L_{rms}, differences needed to exceed 1 dB, as this value estimated the just-noticeable difference (JND) for amplitude (Zwicker & Fastl, 1999). Spectrum-related JNDs for formant frequencies or spectral centroid amount to about 15 Hz for the frequency range in question (Kewley-Port & Watson, 1994; Kendall & Carterette, 1996), whereas spectral-envelope variation has also been linked to lowering JNDs for fundamental frequency (Moore & Moore, 2003). As the latter case points to an even more acute discrimination of spectral change, a more liberal threshold of 5 Hz was adopted for the spectral measures (F_{max}, F_{3dB}, SC). This threshold is based on the discrimination threshold of about 1% for fundamental frequency in complex tones (Zwicker & Fastl, 1999; Moore & Moore, 2003) when applied to the investigated main-formant frequencies. For CV, differences below 10% were considered negligible, because even confounding variables could be shown to introduce greater variability

(see Covariates). Lastly, XC differences below 1% (e.g., 0.3% improved temporal coordination) were considered of too little value to be reported. The threshold for XC was expressed in terms of explained variance, i.e., differences between R^2 values.

Covariates.

As the acoustic measures were based on real-life signals, they may have contained some differences between factor levels that were unrelated to deliberate timbre adjustments by performers. For instance, different rooms typically impose a characteristic *coloration*, i.e., frequency filter, that may induce shifts in the spectral measures. Likewise, the apparent differences in f_0 register between parts likely imposed spectral shifts that lay beyond the performers' control. These possible sources of covariation will therefore be assessed in this section to determine baselines against which to interpret any related effects in the following sections.

The assessment of potential room effects compared fixed reference performances simulated at the *self* positions in the small vs. large rooms. For greater representativeness, this procedure was applied to two selected performances per participant, for parts A and C, yielding 2 × 16 cases. For the median DV measures, the comparison of group medians by room yielded shifts for all spectral measures and L_{rms}: identical horn performances exhibited slightly stronger dynamics in the large than in the small room, with the opposite applying to bassoon. Likewise, the spectral measures varied by about 1% in main-formant frequency between rooms. In terms of CV, the spectral measures exhibited up to 30% more temporal variability in the large room, whereas variability in L_{rms} decreased by up to 10% in the same room. It appears that higher reverberation introduced greater spectral fluctuation, whereas it smoothed out temporal variability in dynamics. As only single performances at the *self* position were considered for the comparison between rooms, the change in XC could not be assessed, because the cross-correlation compared two performers at separate positions. Still, differences in reverberation between rooms may have had an effect on XC as well. As is apparent in Figure 6, the performance at the *other* position (bottom panel) yielded more variability than at the *self* position (top), i.e., signals heard from farther away were also more reverberated. Differences in reverberation between rooms could have therefore modulated the disparity between the two positions, and hence also XC, in some additional way. Unfortunately, these observations suggested that pre-existing, systematic differences between rooms introduced a confounding influence on all measures and across all DVs, compromising the ability to tease apart differences in performer adjustments from those introduced by room acoustics. As a result, obtained ANOVA effects were evaluated against the threshold values quantified above, serving as baselines for the systematic variation. The resulting baselines for median DV between rooms are visualized as the horizontal lines in Figure 7.

[Insert Figure 7 about here]

Spectral covariation with f_0 between parts A, B, and C was quantified on the actual performer data. The comparison considered separate group medians by part, with the spectral shifts expressed relative to part A, which had the highest median f_0. Spectral shifts could also be compared to corresponding changes in f_0 itself, represented by the median across pitches per part, which was weighted by the relative duration of individual pitches. Table 1 displays these comparisons: Although f_0 varied as much as −42%, the spectral shifts were less pronounced, nonetheless exhibiting a monotonic decrease by part, i.e., C was lower than B, which was lower than A. Bassoons exhibited only up to −13% of covariation, whereas horns showed decreases up to −24%. The averaged frequency shifts for B and C were taken as the baselines for spectral shifts induced from f_0 changes alone and are visualized as the horizontal lines in Figure 8.

[Insert Table 1 about here]

[Insert Figure 8 about here]

Given the covariate influence of rooms and f_0, the presentation of results for the factors Room and Interval precedes the three remaining ones. Figures 7, 8, 9, and 11 visualize potential main effects for median DV across all acoustic measures, i.e., F_{max}, F_{3dB}, *SC*, and L_{rms} (individual panels from left to right, respectively). The bars and intervals symbolize means and standard errors, respectively, for within-subject differences between factor levels for the factors Room, Interval, Role, or Phrase. The labels above and below the zero-axis indicate the orientation of a difference between two factor levels. For instance, for the factor Interval (Figure 8) and positive values in *SC*, the spectral centroid was higher for *unison* than *non-unison*; the reverse applies for negative values. In addition, Table 2 summarizes the effects for CV and XC.

Room.

ANOVAs on the median DVs yielded differences between rooms for the spectral measures and sound level that strongly mirrored the expected covariate baselines, as illustrated in Figure 7 by comparing the bars to the corresponding horizontal lines. Assuming these mirrored trends to reflect pre-existing differences in room acoustics, only discrepancies from these baselines beyond the psychoacoustically meaningful threshold will be considered. All but one of the effects fulfilled this criterion, with F_{3dB} for bassoon barely exceeding the baseline by about 5 Hz, $F(1,15) = 22.86$, $p < .01$, $\eta_p^2 = .60$. Also the CV exhibited greater temporal variability in the larger room, as indicated in Table 2. The main-formant measures yielded differences

up to 23%, for both the horn, $F(1,15) \geq 7.74$, $p < .02$, $\eta_p^2 \geq .34$, and the bassoon, $F(1,15) \geq 5.29$, $p < .04$, $\eta_p^2 \geq .26$, which again mirrored the expected trends for room-acoustical variation alone. Similar trends also applied to the temporal coordination, with XC changing up to 8%. Both instruments' L_{rms} exhibited greater XC in the larger room, $F(1,15) \geq 8.32$, $p \leq .01$, $\eta_p^2 \geq .36$. In addition, temporal coordination for horn was also higher in the larger room concerning *SC* and F_{3dB}, $F(1,15) \geq 9.29$, $p < .01$, $\eta_p^2 \geq .38$. In summary, all findings appeared to closely reflect patterns expected from pre-existing, systematic differences in room acoustics and did not allow effects caused by deliberate performer actions to be clearly identified.

[Insert Table 2 about here]

Interval.

The median DV for spectral measures and sound level exhibited higher values in unison than in non-unison. As in the preceding section, the observed differences for Interval generally matched the covariate baselines for f_0 register, as illustrated in Figure 8 when comparing the bars against the horizontal lines. Only for the horn, the spectral measures exhibited higher frequencies for unison, $F(1,15) \geq 106.45$, $p < .01$, $\eta_p^2 \geq .88$, which moreover fell below the baselines by 10 to 20 Hz. These discrepancies could be due to the baselines overestimating the actual within-subjects differences between intervals for hornists, as they were derived from group medians. Nonetheless, these effects could still not be assumed to correspond to blend-rated performer actions, as they were dictated by the musical notation. The pronounced influence of Interval, however, is still important for interpreting interaction effects among the remaining factors.

In addition, as summarized in Table 2, bassoonists showed greater temporal coordination playing in unison than in non-unison, with XC increasing by 4% for εSC, $F(1,15) = 4.82$, $p < .05$, $\eta_p^2 = .24$, although the difference was mainly apparent in the smaller room, Interval × Room: $F(1,15) = 5.69$, $p = .03$, $\eta_p^2 = .28$. By contrast, horns exhibited 8% greater coordination in L_{rms} in non-unison performances, $F(1,15) = 12.00$, $p < .01$, $\eta_p^2 = .44$, with the difference being only half as pronounced in the second phrase, Interval × Phrase: $F(1,15) = 7.76$, $p = .01$, $\eta_p^2 = .34$. These effects were complemented by analogous differences for CV measures of L_{rms}, in that bassoons showed greater temporal variability in unison, $F(1,15) = 4.81$, $p < .05$, $\eta_p^2 = .24$, whereas the opposite applied to horns, $F(1,15) = 6.26$, $p = .02$, $\eta_p^2 = .30$, with the latter being limited to followers, Interval × Role: $F(1,15) = 9.05$, $p < .01$, $\eta_p^2 = .38$. In summary, whereas the Interval factor introduced an upward bias to the acoustical measures for unison performances, which

affected both instruments similarly, the DVs for temporal variability and coordination showed a few opposing trends between instruments.

Role.

The clearest indication for timbre adjustments by performers concerned differences between *leader* and *follower* roles. For the median DVs, role-based differences across spectral features and dynamics become apparent in Figure 9. Musicians produced higher spectral frequencies and increased sound levels as leaders compared to when performing as followers. For bassoon, the main-formant measures were higher for leaders, $F(1,15) \geq 33.02$, $p < .01$, $\eta_p^2 \geq .69$, but this appeared to be limited to non-unison conditions, which was likely related to the f_0 difference between parts B and C, Role × Interval: $F(1,15) \geq 34.76$, $p < .01$, $\eta_p^2 \geq .70$. Likewise, performances for leaders exhibited higher *SC* than did those for followers, $F(1,15) = 60.24$, $p < .01$, $\eta_p^2 = .80$, however, more so in the non-unison conditions, for similar reasons as before, Role × Interval: $F(1,15) = 76.50$, $p < .01$, $\eta_p^2 = .84$. At the same time, L_{rms} increased slightly for leaders, $F(1,15) = 14.49$, $p < .01$, $\eta_p^2 = .49$.

[Insert Figure 9 about here]

The differences obtained for horn exhibited similar patterns. Both F_{max} and F_{3dB} yielded higher frequencies for leaders, $F(1,15) \geq 9.45$, $p < .01$, $\eta_p^2 \geq .39$, with the difference for F_{3dB} appearing to be limited to unison performances, Role × Interval: $F(1,15) = 10.19$, $p < .01$, $\eta_p^2 = .40$. Also *SC* yielded a difference between performer roles, with higher frequencies for leaders, $F(1,15) = 45.91$, $p < .01$, $\eta_p^2 = .75$, being more pronounced for non-unison performances, Role × Interval: $F(1,15) = 6.43$, $p = .02$, $\eta_p^2 = .30$. Analogous differences concerned leaders yielding higher L_{rms}, $F(1,15) = 22.84$, $p < .01$, $\eta_p^2 = .60$, and more so in the non-unison conditions, Role × Interval: $F(1,15) = 30.23$, $p < .01$, $\eta_p^2 = .67$.

In other words, these findings argue that in the attempt to blend with leaders, followers adjusted to 'darker' timbres and, interestingly, spectral features and dynamics changed in a coherent way. For both instruments, *SC* dropped by about 30 Hz and L_{rms} decreased by 1-3 dB for followers. Figure 10 relates the observed differences between performer roles to equivalent spectral-envelope changes. These spectral envelopes (curves) and the indicated acoustic measures (vertical lines) represent medians taken across all performances, collapsed across the remaining factors. Although these aggregate differences do not correspond to within-subject differences, they still show how the effects influenced the entire spectrum. As illustrated by the black arrows traversing the pairs of spectral envelopes, the main formants of followers (dark grey) receded in frequency and level compared to the leaders' (light grey). This was reflected in analogous differences across the acoustic

measures (vertical lines), although the detailed analysis mirrors the observed differences between instruments (e.g., differences in line width). The main formants in unison bassoon performances remained fixed (top-left panel), whereas the change in *SC* suggested spectral adjustments relative to the main formant, which co-occurred with a slight change in L_{rms}. For the same unison conditions, the horns exhibited more change in formant measures and sound level (top-right).

[Insert Figure 10 about here]

With regard to temporal variation, the DVs quantifying the CV exhibited instrument-specific effects, as summarized in Table 2. Leading hornists varied more than followers along F_{3dB} and *SC*, $F(1,15) \geq 9.15$, $p < .01$, $\eta_p^2 \geq .38$, whereas the contrary applied to bassoonists across all spectral measures, $F(1,15) \geq 22.42$, $p < .01$, $\eta_p^2 \geq .60$. For both instruments, these effects were limited to non-unison performances, which suggests that they arose from instrument-specific issues related to parts B and C, Role × Interval: $F(1,15) \geq 5.93$, $p < .03$, $\eta_p^2 \geq .28$. For instance, the low registral range of part C posed more playing difficulty to hornists than to bassoonists. Other role-dependent differences were specific to horns, in which temporal variation of L_{rms} was greater for followers, $F(1,15) = 17.07$, $p < .01$, $\eta_p^2 = .53$, whereas the temporal coordination as quantified by XC was up to 3% higher for leaders concerning εF_{3dB} and εSC, $F(1,15) \geq 5.68$, $p \leq .03$, $\eta_p^2 \geq .28$. In summary, the effects between performer roles for temporal variation and coordination yielded less coherent patterns than those for median DVs. The observed tendencies were mainly instrument-specific, which seemed more pronounced for spectral variation in the lower pitch registers.

Phrase.

Comparisons between the first and second phrases indicated that both musicians adapted their playing throughout a performance, adjusting their timbres toward an assumedly improved blend. With regard to median DV, leading bassoonists lowered *SC* by about 12 Hz towards the second phrase, whereas followers increased by 10 Hz, still remaining below leaders, Phrase × Role: $F(1,15) = 25.63$, $p < .01$, $\eta_p^2 = .63$. The effect for followers appeared limited to non-unison conditions, whereas in unison, followers did not vary *SC* in their performances, Phrase × Role × Interval: $F(1,15) = 31.22$, $p < .01$, $\eta_p^2 = .68$. This notable interaction revealed that even leaders attempted to close larger gaps in *SC*, whereas followers fulfilled the same objective by remaining stable or closing gaps in the opposite direction. Hornists showed similar effects, although without interactions with other factors, as illustrated in Figure 11. The formant measures decreased by about 5 Hz in the second phrase, $F(1,15) \geq 6.69$, $p \leq .02$, $\eta_p^2 \geq .31$. Likewise, L_{rms} also decreased by about 1 dB throughout performances, $F(1,15) = 28.22$, $p < .01$, $\eta_p^2 \geq .65$.

Overall, the difference in spectral frequencies between phrases spanned between 5 Hz and 12 Hz, which given the prior discussion of thresholds may not have yielded clearly perceptible differences in all cases.

[Insert Figure 11 about here]

Similar effects for temporal coordination supported the previous findings, as outlined in Table 2. For L_{rms}, the second phrase yielded 6% and 8% higher XC for bassoon, $F(1,15) = 37.93$, $p < .01$, $\eta_p^2 = .72$, and horn, $F(1,15) = 125.05$, $p < .01$, $\eta_p^2 = .89$, respectively. Similarly, the coordination in εSC also increased in the later phrase by 3% for bassoon, $F(1,15) = 9.86$, $p < .01$, $\eta_p^2 = .40$, and 5% for horn, $F(1,15) = 19.14$, $p < .01$, $\eta_p^2 = .56$. The increased coordination in the later phrase may have been related to the notated *crescendo-decrescendo* (see Figure 3, measures 13-14), which could likewise have explained a corresponding increase in horn players' CV for L_{rms}, $F(1,15) = 92.41$, $p < .01$, $\eta_p^2 = .86$. In summary, performances in the second phrase exhibited a greater degree of temporal coordination, and they also seemed to involve adjustments toward a more similar and moderately 'darker' spectrum.

Communication.

Among the acoustic measures, no clear indications were obtained that the absence of auditory feedback from the follower affected performances differently than in the unimpaired case. Of the few statistically significant findings, all fell below the pre-defined thresholds for psychoacoustically meaningful differences.

Discussion

When two musicians aim to achieve a blended timbre during performance, they coordinate their playing in a certain way. Both performers aim for the idealized timbre the musical score conveys, which usually also implies the instrument that should lead in performance. The leading musician determines timing, intonation, and phrasing, providing reference cues that accompanying musicians closely follow, who likely also adjust their timbres to ensure blend. The employed strategies of performer coordination may or may not be influenced by whether they are playing in unison or non-unison, whether they perform in different venues, or whether the leading instrument is unable to hear the other musician (as in offstage playing, for example). These factors were studied for pairs of one bassoon and one horn player, focusing on the timbral adjustments they employed. Performances were evaluated over their time courses through a set of acoustic measures, complemented by self-assessment from the performers, delivering a differentiated picture of how performers adjust timbre in achieving blend.

Measuring timbre adjustments as they occur in the realistic setting of musical performance involves a high degree of complexity. These

adjustments were evaluated through spectral features, which in some cases, however, seemed inseparable from covariation with pitch and dynamics. These covariates are what a musical score essentially communicates to performers and although timbre is implied through instrumentation and articulation markings, for a given instrument it also occurs as a by-product of notated pitches and dynamics. These covariates also determine how performers excite their instruments' acoustic systems, in turn establishing inherent links to the resulting spectral properties. Although correlation analyses on their own do not prove causal relationships, the inherent coupling of pitch, dynamics, and spectral properties in wind instruments has been established physically (Benade, 1976), and this should hence justify their association.

Correlations between spectral measures and the covariates of pitch (f_0) and dynamics (L_{rms}) are visualized in Figure 12. As individual differences across performers and their instruments were to be expected, the evaluation considered correlations across all performances of individual players and then summarized these as medians and interquartile ranges for bassoon and horn separately. An impact of pitch variation becomes clearly apparent, reflected in positive correlations, $r \approx .55$, between f_0 and all spectral measures. This applied to both instruments, with spectral centroid (*SC*) being most affected and there being little variability among players. The potential influence of dynamics on the spectral measures differed fundamentally across instruments. Whereas correlations with L_{rms} for bassoonists were nearly absent, $r \approx -.10$, hornists exhibited clearly positive correlations, $r \approx .40$. In addition, there was also a trend for positive correlations between pitch and dynamics, which differed in magnitude between instruments. In summary, pitch appears to induce substantial spectral change, and due to it being dictated by musical notation, these changes lie beyond performers' expressive control.

Although the results from the experiment showed tendencies for increases in sound level to reflect increases in spectral measures (F_{max}, F_{3dB}, *SC*), a linear covariation was only obtained for horns. Regardless of these differences, dynamics may still have afforded performers of both instruments greater liberty in timbral control, although not necessarily in the same way. Subtle changes in dynamics that remain within the notated dynamic markings could thus be used for slight timbre adjustments and may be more easily achieved than adjustments independent of both dynamics and pitch. Experienced orchestrators likely have internalized the inherent links between pitch, dynamics, and timbral properties in their instrumentation knowledge (e.g., *pitch-driven dynamics*), whereas the current findings argue that research on timbre perception that aims to situate it within musical practice should abandon its definition as that residual quality alongside pitch and dynamics, instead accepting the notion of it being closely entwined with the other musical parameters (McAdams & Goodchild, in press).

[Insert Figure 12 about here]

Assigning roles to performers yielded the clearest effects for timbral adjustments related to blend. Players acting as leaders indeed functioned as a reference toward which followers oriented their playing. In order to achieve blend, followers adjusted towards *darker* timbres compared to when they performed as leaders. For both instruments, the darker timbre corresponded to shifts of *SC* by about 30 Hz towards lower frequencies, whereas the main formant shifted as well, but only for the horn. These selective spectral adjustments can be compared with similar strategies undertaken by singers to blend into a choir (Goodwin, 1980; Ternström, 2003). At the same time, a *darker* timbre occurred together with *softer* dynamics, which suggests that performers may have partially achieved the timbre change through subtle changes in dynamics, in addition to potential changes in embouchure or the position of the right hand in the bell of the horn. The extent to which spectral change was employed varied between instruments, with the horn clearly producing more change — it is also known to be the timbrally more versatile instrument. Due to the nature of the within-subjects design, these role comparisons considered how the same musicians performed differently as followers than as leaders, i.e., they did not assess how bassoonist followers darkened their timbre relative to hornist leaders and vice versa. At the least, Figure 10 suggests that as followers, hornists lowered their upper-bound formant frequencies (F_{3dB}) to be about the same as that of the bassoonists, which is necessary to avoid a marked decrease in perceived blend (Lembke & McAdams, 2015).

With regard to the magnitude of changes in dynamics, differences in L_{rms} (e.g., 1-3 dB) were not so pronounced as to signify a departure from the notated dynamic marking *piano*. From interviewing players of both instruments, musicians appear to consciously consider adjustments of both dynamics and timbre as strategies to achieve blend. For instance, in accompanying a leading instrument, a hornist described his goal as achieving a "rounder" or less brilliant timbre, at the same time reporting that playing with woodwinds, he would need to avoid "overpowering" the other instrument in dynamics. Likewise, a bassoonist reported the importance of loudness balance to blend, also clarifying that to her, dynamics and timbre were not independent. Yet, it cannot be ruled out that spectral changes occurred as by-products of sound-level adjustments made for wholly other reasons than achieving blend. For instance, adjusting the *self-to-other ratio*, i.e., the sound-level difference between oneself and another performer, may improve communication amongst musicians (Ternström, 2003; Keller, 2014; Fulford et al., 2014). Despite this possibly confounding influence, it still seems justified to assume some quantity of the observed spectral changes to stem from blend-related adjustments, as no clear correlation between L_{rms} and the spectral measures is apparent, especially for the bassoon (see Figure 12).

Unison performances were indeed perceived as yielding significantly higher blend than their non-unison counterparts, but the mean difference between the two was merely 4% of the full range of the rating scale. This

small difference may be explained in a number of ways. Listening experiments conducted in the past obtained clearer differences in blend ratings between unison and non-unison. In the current experiment, however, participants provided retrospective ratings alongside the more demanding performance task, with the ratings also being well separated in time, which did not allow immediate comparisons of unison vs. non-unison performances. Furthermore, performers were asked to use the rating scale based on their previous musical experience, i.e., judging performances and blend relative to what they had learned was achievable in musical practice. In addition, blending could have also been understood as how 'coupled' the musicians' performance was, i.e., related to additional factors such as synchrony and intonation. Lastly, the musicians' own playing could have partially masked their perception of the other player (e.g., hearing their instrument in greater proximity and via bone conduction), which does not compare to conventional listening experiments, where participants are presented a comparatively balanced rendition of two instruments. Together, these factors could have led to the less pronounced rating differences between the interval conditions.

Nonetheless, higher blend may still relate to unison performances influencing player coordination more critically. In unison, the performance ratings suggest that followers gave higher ratings than did leaders, which could imply that leaders were generally less satisfied with their performance, given their more important role and responsibility for its success. By contrast, non-unison performances yielded higher ratings by leaders than followers. This result could be related to part C being located in a low register, which may have led to some noticeable playing difficulty for a few players. While communication directivity did not appear to affect performances as measured acoustically, the only time it did become relevant concerned unison performances, as impaired communication was judged to be detrimental to musicians' performance. In a similar way, although room-acoustical effects related to blend could not be deduced from the acoustic measures, the performance ratings revealed more pronounced effects between performer roles in the smaller, less reverberant room. These effects suggest that performer coordination between instruments was more critical in the smaller room, which may have allowed more subtle differences to become audible. Indeed, temporal coordination for one spectral measure was found to be higher for unison and the smaller room, although this remained limited to global spectral change (εSC) and bassoons.

Several indications suggest that musicians improved their coordination throughout a performance. The temporal coordination for both instruments improved in the later phrase for both dynamics (L_{rms}) and global spectral change (εSC) by up to 5% and 8%, respectively. It should be noted that while median values of temporal coordination (XC) across both measures and instruments were comparable, $r \approx .24$, they indicate a fairly weak positive correlation, which suggests that timbre-related performer coordination does not operate at a fine time resolution but only appears to apply to larger time

segments, such as first vs. second phrase. Furthermore, the assessment of temporal change suggests that even leaders adjust their timbre. For instance, regardless of assigned role, horn players slightly reduced their main-formant frequencies and dynamic level in the second phrase. Although these changes were of considerably smaller magnitude than the ones between performer roles (e.g., 5 vs. 30 Hz), and likely of lesser perceptual salience, performer coordination appears to motivate adjustments by both musicians to a limited degree. Overall, this result both suggests that performer coordination adapts over time, ideally leading to an improvement, and that the reference function of leaders still allows for a certain degree of bilateral adjustment between performers. As there was no indication that performer coordination was modulated by either communication impairment or performance venue among the acoustic measures, the strategies musicians employ in achieving blend appear to be fairly robust to acoustical factors.

This investigation represents a case study by featuring two instruments that commonly form a blended timbre in the orchestral literature. Given the high timbral similarity between bassoon and horn, an effect of performer roles was obtained across both instruments, i.e., regardless of which was leading in performance, whereas obtaining a role-based effect would become less likely when there are starker differences between instrument timbres. In the latter scenario, the more dominant timbre would seem predisposed to assume the lead and serve as the reference, into which the other instrument would either succeed or fail to blend. This case concerns what Sandell (1995) referred to as the *augmented timbre*, in which a dominant instrument is timbrally enriched by another instrument. With this case being a common goal in orchestration, its success depends on the ability of the other instrument to blend into the context defined by the reference. Either its spectral envelope lacks any prominent features that would otherwise 'challenge' the dominant instrument or it bears a sufficiently high resemblance to the latter. In the current investigation, both instrument timbres were similar, yet, the greater timbral versatility of the horn allowed it to blend into a bassoon sound (see Figure 10), whereas the bassoon would not have succeeded in adjusting towards a more brilliant or 'brassy' timbre in return. This imbalance in timbral adjustments, paired with instrument-specific issues related to the playability of parts, could explain the differences in performance ratings between instruments. For example, hornists generally gave higher ratings of their performances as leaders than as followers, which could be linked to the greater ease of playing in their default timbre as leaders, as opposed to having to adjust to a substantially darker timbre as followers. This implies that even in this common pairing, the horn may generally assume the more dominant role over bassoons, which also manifests itself in the orchestral repertoire. Their combination in unison is in fact less common, likely explained by their high similarity not adding much timbral enrichment, whereas their combination in non-unison is widespread. In the latter cases, bassoons are often substituted for missing horns, because up to the mid-nineteenth century,

orchestras generally only included two horns. The addition of bassoons overcame this limitation, as is also the case in the investigated orchestral passage by Mendelssohn-Bartholdy. In practice, bassoonists more often find themselves blending into the horn timbre than vice versa.

Despite the various scenarios concerning instrument combinations as well as dominance or role relationships, a common rule seems to apply to all: In attaining perceptual blend, the accompanying instrument darkens its timbre in order to avoid 'outshining' the leading, dominant instrument. In other words, when an accompanying instrument blends into the leading instrument, it adopts a strategy of remaining subdued and low-key, very similar to how it subordinates itself to the lead instrument's cues for intonation, timing, and phrasing.

Conclusion

The current investigation showcases how the orchestration goal of achieving blended timbres is mediated by factors related to musical performance. For instrument combinations exhibiting similar timbres (e.g., bassoon and horn), the assignment of performer roles may determine which instrument serves as a reference toward which accompanying musicians adapt their timbre to be darker. In an arbitrary combination of instruments, a possible dominance of one timbre likely biases that instrument toward assuming the reference and leading role, requiring that another instrument be able to blend in, otherwise resulting in a heterogeneous timbre. With respect to previous research on musical performance, the current findings illustrate a case in which performer coordination, as related to concepts like *joint action* and *leadership*, directly applies to performers' control of timbre. Achieving a blended timbre requires coordinated action in which an orchestrator's intention becomes the common aim of two or more performers, involving strategies based on relative performer roles that ensure the idealized goal is realized. Standing in the limelight of performance, leading musicians assume the responsibility over the accurate and expressive delivery of musical ideas, whereas the accompanist's primary concern is to blend in, and if successful, remain somewhat obscured in the lead instrument's timbral shadow.

References

Benade, A. H. (1976). *Fundamentals of musical acoustics*. New York: Oxford University Press.

Bregman, A. S. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: MIT Press.

D'Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., Aloimonos, Y., & Fadiga, L. (2012). Leadership in orchestra emerges from the causal relationships of movement kinematics. *PloS one*, *7*(5), e35757.

Doval, B. & Rodet, X. (1991). Estimation of fundamental frequency of musical sound signals. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toronto (pp. V–3657–V–3660).

Fabiani, M. & Friberg, A. (2011). Influence of pitch, loudness, and timbre on the perception of instrument dynamics. *Journal of the Acoustical Society of America*, *130*(4), EL193–199.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fulford, R., Hopkins, C., Seiffert, G., & Ginsborg, J. (2014). Sight, sound and synchrony: effects of attenuating auditory feedback on duo violinists' behaviours in performance. In *Proc. 13th International Conference on Music Perception and Cognition (ICMPC)*, Seoul (pp. 166–171).

Goad, P. J. & Keefe, D. H. (1992). Timbre discrimination of musical instruments in a concert hall. *Music Perception*, *10*(1), 43–62.

Goebl, W. & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, *26*(5), 427–438.

Goodman, E. (2002). Ensemble performance. In J. Rink (Ed.), *Musical performance: A guide to understanding* (pp. 153–167). Cambridge: Cambridge University Press.

Goodwin, A. W. (1980). An acoustical study of individual voices in choral blend. *Journal of Research in Music Education*, *28*(2), 119–128.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. *Journal of Educational and Behavioral Statistics*, *17*(4), 315–339.

Keller, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity* (pp. 205–221). Amsterdam: IOS Press.

Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 260–282). Oxford: Oxford University Press.

Keller, P. E. & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception*, *28*(1), 27–46.

Kendall, R. A. & Carterette, E. C. (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception*, *8*(4), 369–404.

Kendall, R. A. & Carterette, E. C. (1993). Identification and blend of timbres as a basis for orchestration. *Contemporary Music Review*, *9*(1), 51–67.

Kendall, R. A. & Carterette, E. C. (1996). Difference thresholds for timbre related to spectral centroid. In *Proc. 4th International Conference on Music Perception and Cognition (ICMPC)*, Montreal (pp. 166–171).

Kewley-Port, D. & Watson, C. S. (1994). Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America*, *95*(1), 485–496.

Koechlin, C. (1954). *Traité de l'orchestration : en quatre volumes*. Paris: M. Eschig.

Lembke, S.-A. & McAdams, S. (2015). The role of spectral-envelope characteristics in perceptual blending of wind-instrument sounds. *Acta Acustica united with Acustica*, *101*(5), 1039–1051.

McAdams, S. & Goodchild, M. (in press). Musical structure: Sound and timbre. In R. Ashley & R. Timmers (Eds.), *Routledge companion to music cognition*. New York: Routledge.

Meyer, J. (2009). *Acoustics and the performance of music* (5th ed.). New York: Springer.

Moore, B. C. & Moore, G. A. (2003). Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects. *Hearing Research*, *182*(1-2), 153–163.

Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception & Psychophysics*, *41*(6), 525–533.

Papiotis, P., Marchini, M., Perez-Carrillo, A., & Maestre, E. (2014). Measuring ensemble interdependence in a string quartet through analysis of multidimensional performance data. *Frontiers in Psychology*, *5*, 963.

Paul, S. (2009). Binaural recording technology: A historical review and possible future developments. *Acta Acustica united with Acustica*, *95*(5), 767–788.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: extracting audio descriptors from musical signals. *Journal of the Acoustical Society of America*, *130*(5), 2902–2916.

Rakowski, A. (1990). Intonation variants of musical intervals in isolation and in musical contexts. *Psychology of Music*, *18*(1), 60–72.

Rasch, R. A. (1988). Timing and synchronization in ensemble performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 70–90). Oxford: Clarendon Press.

Reuter, C. (1996). *Die auditive Diskrimination von Orchesterinstrumenten - Verschmelzung und Heraushörbarkeit von Instrumentalklangfarben im Ensemblespiel (The auditory discrimination of orchestral instruments: Fusion and distinguishability of instrumental timbres in ensemble playing)*. Frankfurt am Main: P. Lang.

Rimsky-Korsakov, N. (1964). *Principles of orchestration*. New York: Dover Publications.

Rodet, X. & Schwarz, D. (2007). Spectral envelopes and additive + residual analysis/synthesis. In J. W. Beauchamp (Ed.), *Analysis, synthesis, and perception of musical sounds* (pp. 175–227). New York: Springer.

Sandell, G. J. (1991). *Concurrent timbres in orchestration: A perceptual study of factors determining blend*. PhD thesis, Northwestern University.

Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration. *Music Perception*, *13*(2), 209–246.

Tardieu, D. & McAdams, S. (2012). Perception of dyads of impulsive and sustained instrument sounds. *Music Perception*, *30*(2), 117–128.

Ternström, S. (2003). Choir acoustics–an overview of scientific research published to date. *International Journal of Research in Choral Singing*, *1*(1), 3–12.

Villavicencio, F., Röbel, A., & Rodet, X. (2006). Improving LPC spectral envelope extraction of voiced speech by true-envelope estimation. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse (pp. I–869–I–872).

Zwicker, E. & Fastl, H. (1999). *Psychoacoustics: Facts and models* (2nd ed.). Berlin: Springer.

Author Note

The authors would like to acknowledge CIRMMT for student awards to Scott Levine and Sven-Amin Lembke and for the use of its equipment and facilities. We would like to expressly thank its technical staff, Harold Kilianski, Yves Méthot, and Julien Boissinot, for their technical advice and assistance. The authors also thank Martha de Francisco from the Sound Recording program at the Schulich School of Music, McGill University for her co-supervision and valuable feedback throughout the project. This research was partly funded by a Canadian National Sciences and Engineering Research Council grant (RGPIN 312774-2010) and a Canada Research Chair to Stephen McAdams. We thank three anonymous reviewers for valuable comments on earlier versions of this article.

Correspondence concerning this article should be directed to Sven-Amin Lembke, currently at De Montfort University, Clephan Building, Room 00.07b, Leicester LE1 9BH, United Kingdom; email: sven-amin.lembke@dmu.ac.uk.

Tables

Table 1: Influence of pitch differences (f_0) among musical parts on the spectral measures (F_{max}, F_{3dB}, SC).

Part	f_0		Bassoon			Horn		
(rel. to A)	Hz	%	F_{max}	F_{3dB}	SC	F_{max}	F_{3dB}	SC
B	-62	-25	-4	-2	-6	-19	-12	-13
C	-104	-42	-13	-7	-13	-24	-12	-21

Note. The covariation was evaluated for parts B and C relative to A (in % if not indicated otherwise), quantified as medians across all performances of a part. f_0 per part considered the median across the pitches of all performed notes, weighted by their relative durations.

Table 2: Summary table of main effects for DVs evaluating performers' temporal variability (CV) and temporal coordination (XC) across acoustic measures for the factors Room, Interval, Role, and Phrase.

			Bassoon				**Horn**			
			F_{max}	F_{3dB}	SC	L_{rms}	F_{max}	F_{3dB}	SC	L_{rms}
	Room									
CV	more variability	larger	■	■	▒	▒	■	■	▒	▒
		smaller	□	□	▒	▒	□	□	▒	▒
XC	more coordination	larger	▒	▒	▒	■	▒	■	■	■
		smaller	▒	▒	▒	□	▒	□	□	□
	Interval									
CV	more variability	unison	▒	▒	▒	■	▒	▒	▒	□
		non-unison	▒	▒	▒	□	▒	▒	▒	■
XC	more coordination	unison	▒	▒	■	▒	▒	▒	▒	□
		non-unison	▒	▒	□	▒	▒	▒	▒	■
	Role									
CV	more variability	leader	□	□	□	▒	▒	■	■	□
		follower	■	■	■	▒	▒	□	□	■
XC	more coordination	leader	▒	▒	▒	▒	▒	■	■	▒
		follower	▒	▒	▒	▒	▒	□	□	▒
	Phrase									
CV	more variability	phrase 1	▒	▒	▒	▒	▒	▒	▒	□
		phrase 2	▒	▒	▒	▒	▒	▒	▒	■
XC	more coordination	phrase 1	▒	▒	□	□	▒	▒	□	□
		phrase 2	▒	▒	■	■	▒	▒	■	■

Note. Vertically adjacent pairs of black and white fields represent main effects and their orientation. For instance, in the top row for bassoon and F_{max}, more temporal variability was obtained in the larger room (black) than in the smaller room (white). No significant differences were found for the grey-shaded fields.

Figure Captions

Figure 1: Spectral envelopes for bassoon (white area) and horn (grey area) at dynamic marking *piano*, estimated using the pitch-generalized method (Lembke & McAdams, 2015). Frequency descriptors for the main formant, F_{max} (solid red) and F_{3dB} (dashed red), and the global spectral centroid *SC* (solid blue) reflect the similarity in prominent spectral-envelope features of the two instruments. The spectral envelopes are offset vertically by 6 dB for better comparison.

Figure 2: Spectrogram of horn playing an A-major scale from A2 to A4, based on time-variant spectral-envelope estimates (*True Envelope*, Villavicencio et al., 2006). The plot displays spectral-envelope magnitude (colormap at the far right) along frequency (y-axis) and time (x-axis), spectral measures F_{max}, F_{3dB}, and *SC* and fundamental frequency f_0 (solid red, dashed red, blue, and white curves, respectively) as well as sound level L_{rms} summed across frequencies (separate horizontal strip at the bottom). Sound levels were normalized to the maximum level of the excerpt (0 dB).

Figure 3: Musical parts A, B, and C in A-major transposition, based on Mendelssohn-Bartholdy's *A Midsummer Night's Dream*. The 'V' marks the separation into the first and second phrases (see *Musical factors*).

Figure 4: Within-subject differences in performance ratings across the factor interactions Role × Interval (top; leader minus follower) and Communication × Interval (bottom; one-way minus two-way). Bars and intervals represent means and standard errors, respectively.

Figure 5: Within-subject differences in performance ratings across the factor interactions Instrument × Role (top; bassoon minus horn) and Role × Instrument × Room (bottom; leader minus follower). Bars and intervals represent means and standard errors, respectively.

Figure 6: Spectrograms of a joint performance of the unison part by one bassoon (top) and one horn (bottom) player, employing TE estimation (Villavicencio et al., 2006). Curves display the time series of (smoothed) acoustic measures F_{max}, F_{3dB}, *SC* and f_0; the separate horizontal strip at the bottom displays L_{rms}. See caption of Figure 2 for further details.

Figure 7: Within-subject differences for the Room factor (large minus small) and DVs evaluating performers' medians of the acoustic time series derived from the performances (individual panels) by instrument. Labels above and below zero indicate the orientation of differences between the two factor levels. For example, positive values for L_{rms} signify that the sound level was higher while playing in the *large* than in the *small* room. Bars and intervals represent means and standard errors, respectively; asterisks (*) indicate significant main effects. Black horizontal lines indicate the expected covariation arising from room-acoustical variability alone.

Figure 8: Within-subject differences for the Interval factor (unison minus non-unison) and DVs evaluating performers' medians of the acoustic time series derived from the performances (individual panels) by instrument. Black horizontal lines indicate the expected covariation arising from f_0-register variability alone. See Figure 7 caption.

Figure 9: Within-subject differences for the Role factor (leader minus follower) and DVs evaluating performers' medians of the acoustic time series derived from the performances (individual panels) by instrument. See Figure 7 caption.

Figure 10: Spectral-envelope change as a function of performer roles (curves), by Interval (top and bottom panels) and instrument (left and right panels). Arrows trace the adjustments toward lower frequencies and sound levels from *leader* to *follower* roles. Corresponding shifts along median DV for F_{max}, F_{3dB}, and SC (vertical lines) illustrate the trend toward a 'darker' spectrum, with the line width corresponding to the actual shift in frequency.

Figure 11: Within-subject differences for the Phrase factor (1st phrase minus 2nd phrase) and DVs evaluating performers' medians of the acoustic time series derived from the performances (individual panels) by instrument. See Figure 7 caption.

Figure 12: Quantification of the covariation introduced by pitch (f_0) and dynamics (L_{rms}) on the spectral measures (F_{max}, F_{3dB}, SC). Bar heights and error bars represent medians and interquartile ranges, respectively, of Pearson correlation coefficients computed per performer, which considered all available time-series data ($N \approx 65{,}000$).

Fig. 1

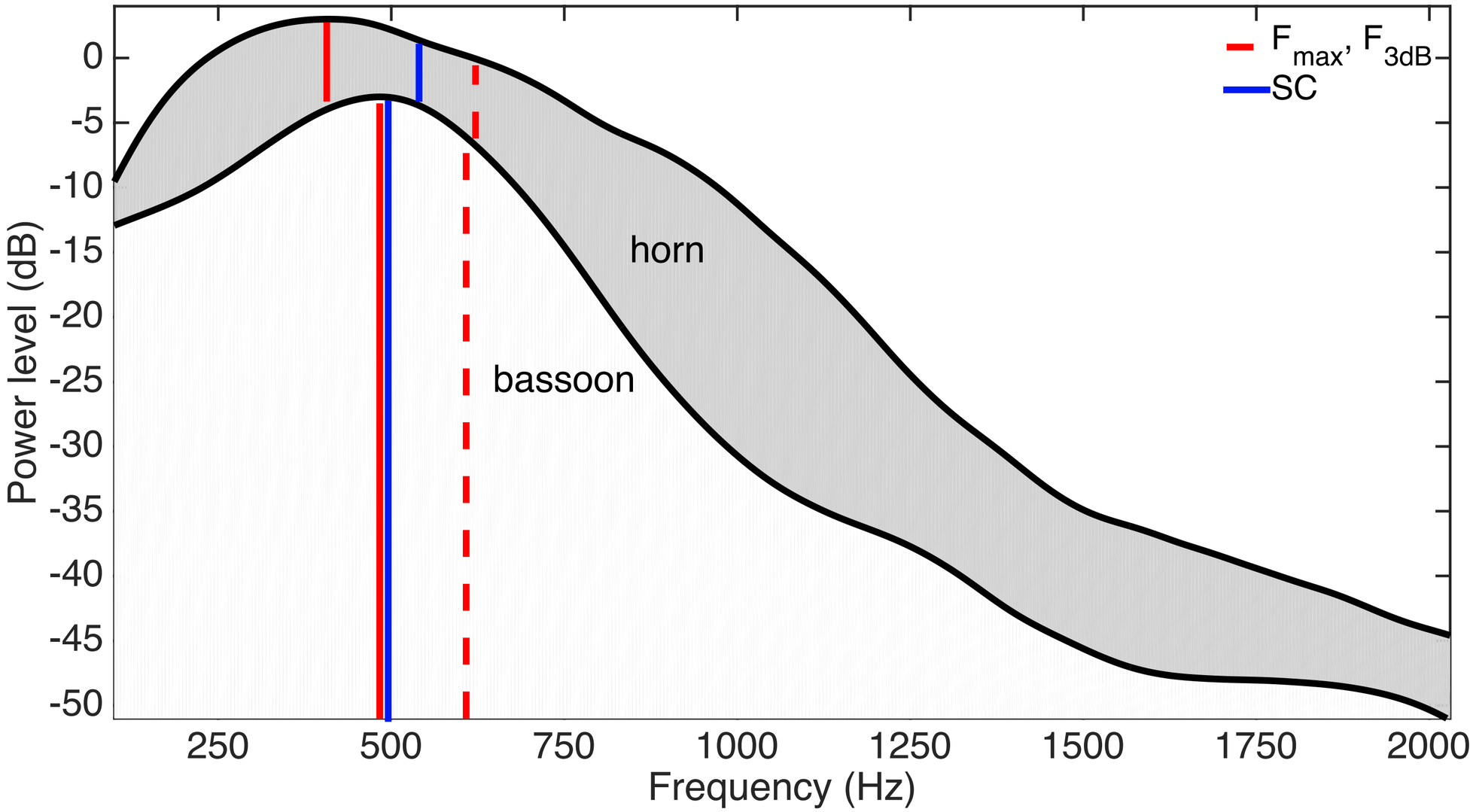
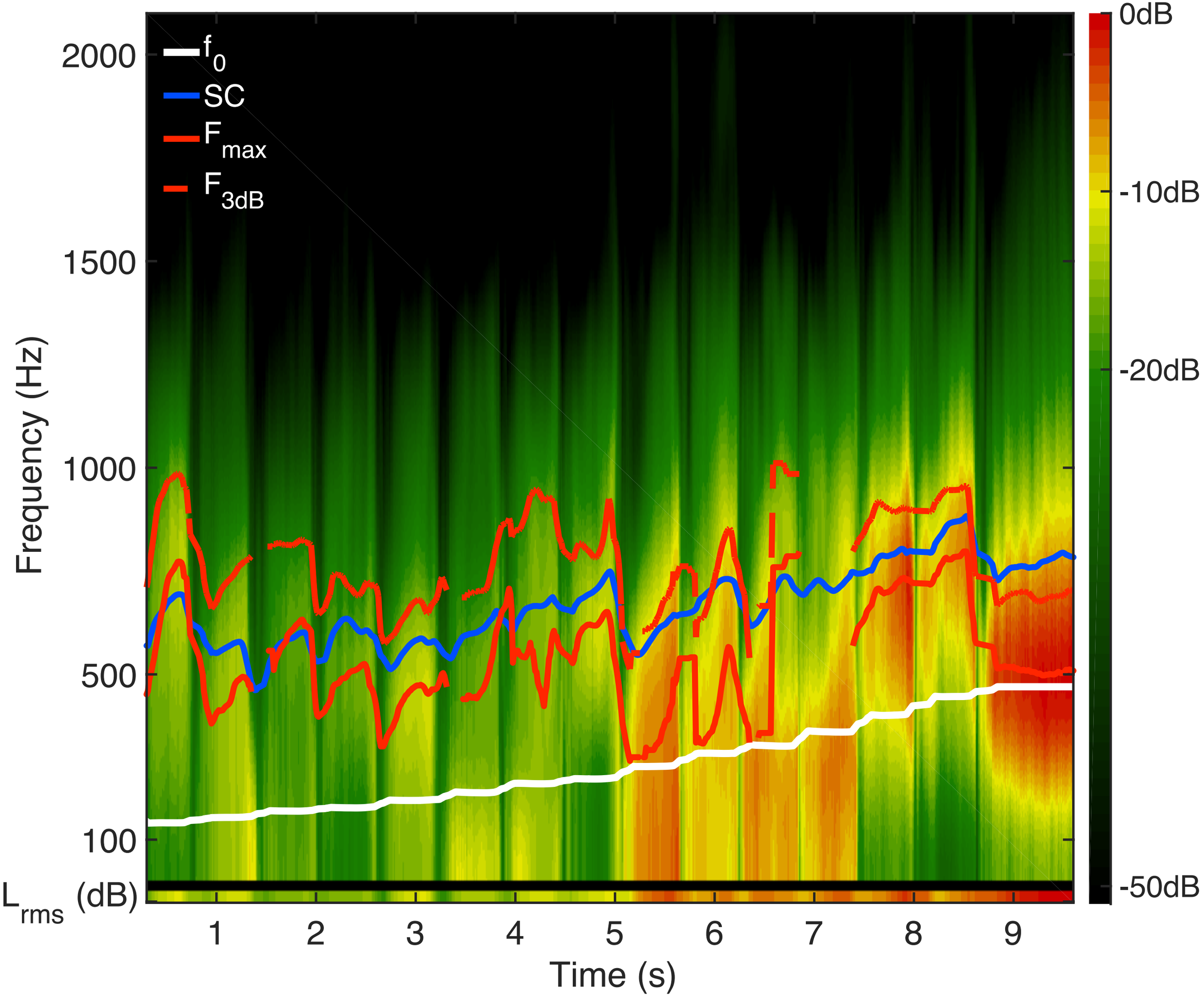

F max, F 3dB
SC
horn
bassoon
Power level (dB)
0
-5
-10
-15
-20
-25
-30
-35
-40
-45
-50
250
500
750
1000
1250
1500
1750
2000
Frequency (Hz)

Fig. 2


f0
SC
Fmax
F3dB
2000
1500
1000
500
100
Frequency (Hz)
0dB
-10dB
-20dB
-50dB
Lrms (dB)
1
2
3
4
5
6
7
8
9
Time (s)

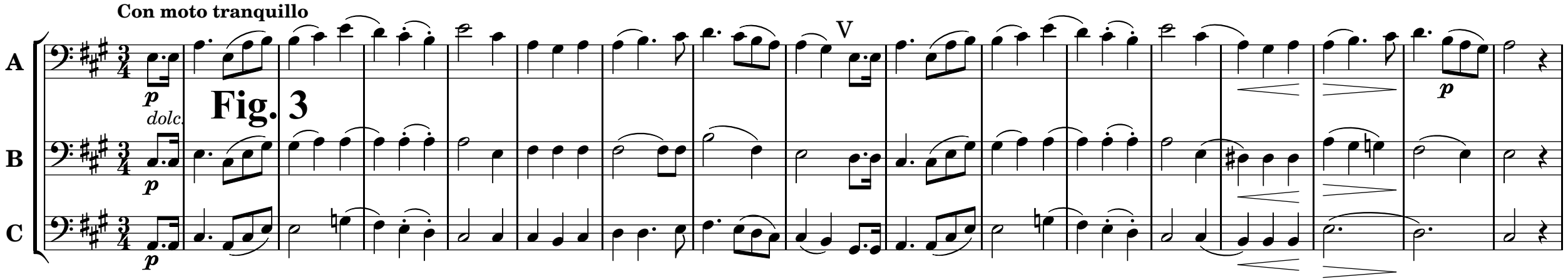

Con moto tranquillo
Fig. 3
dolc.
V
A
B
C

Fig. 4

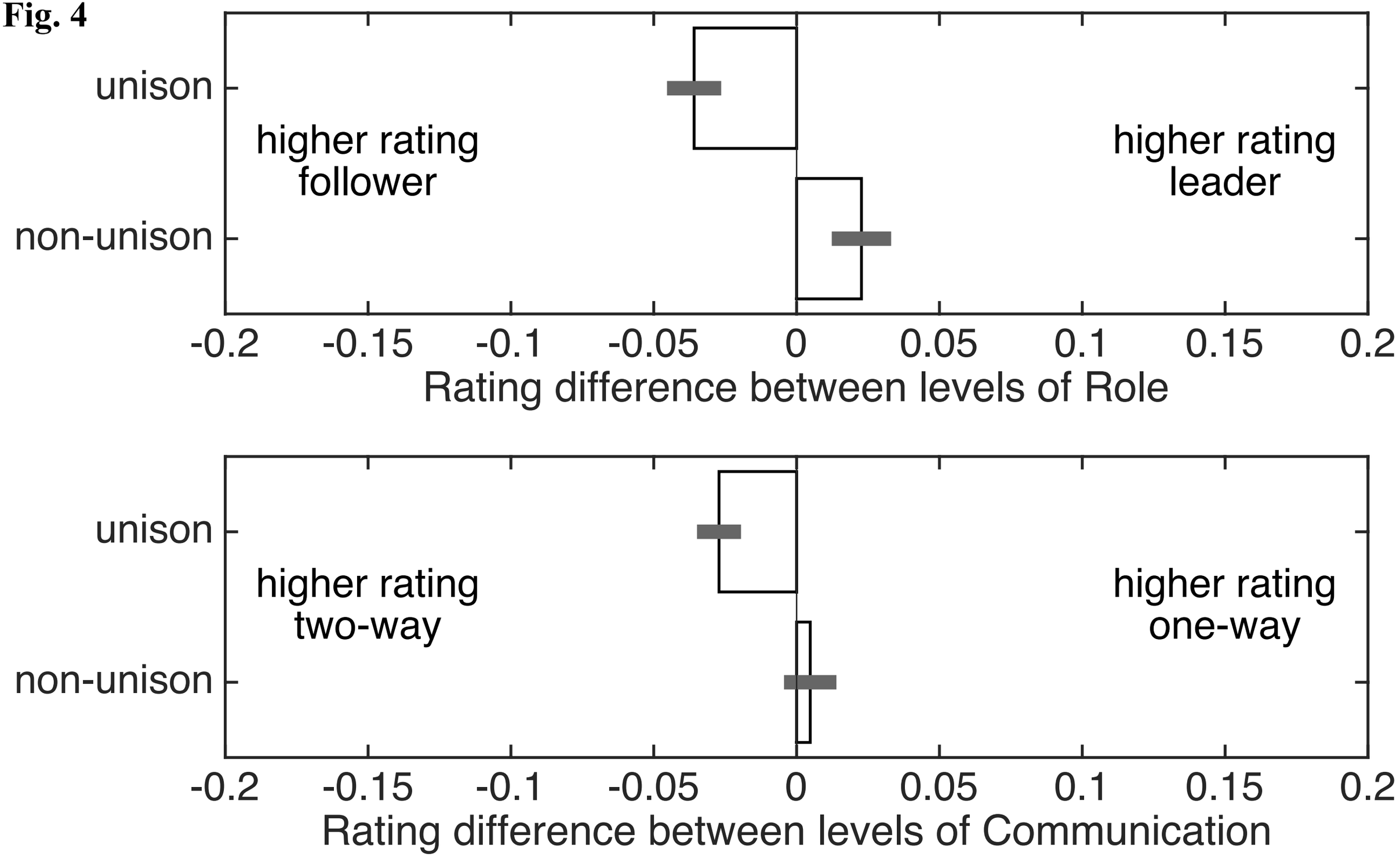

unison
non-unison
higher rating
follower
higher rating
leader
-0.2
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
0.2
Rating difference between levels of Role
unison
non-unison
higher rating
two-way
higher rating
one-way
-0.2
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
0.2
Rating difference between levels of Communication

Fig. 5

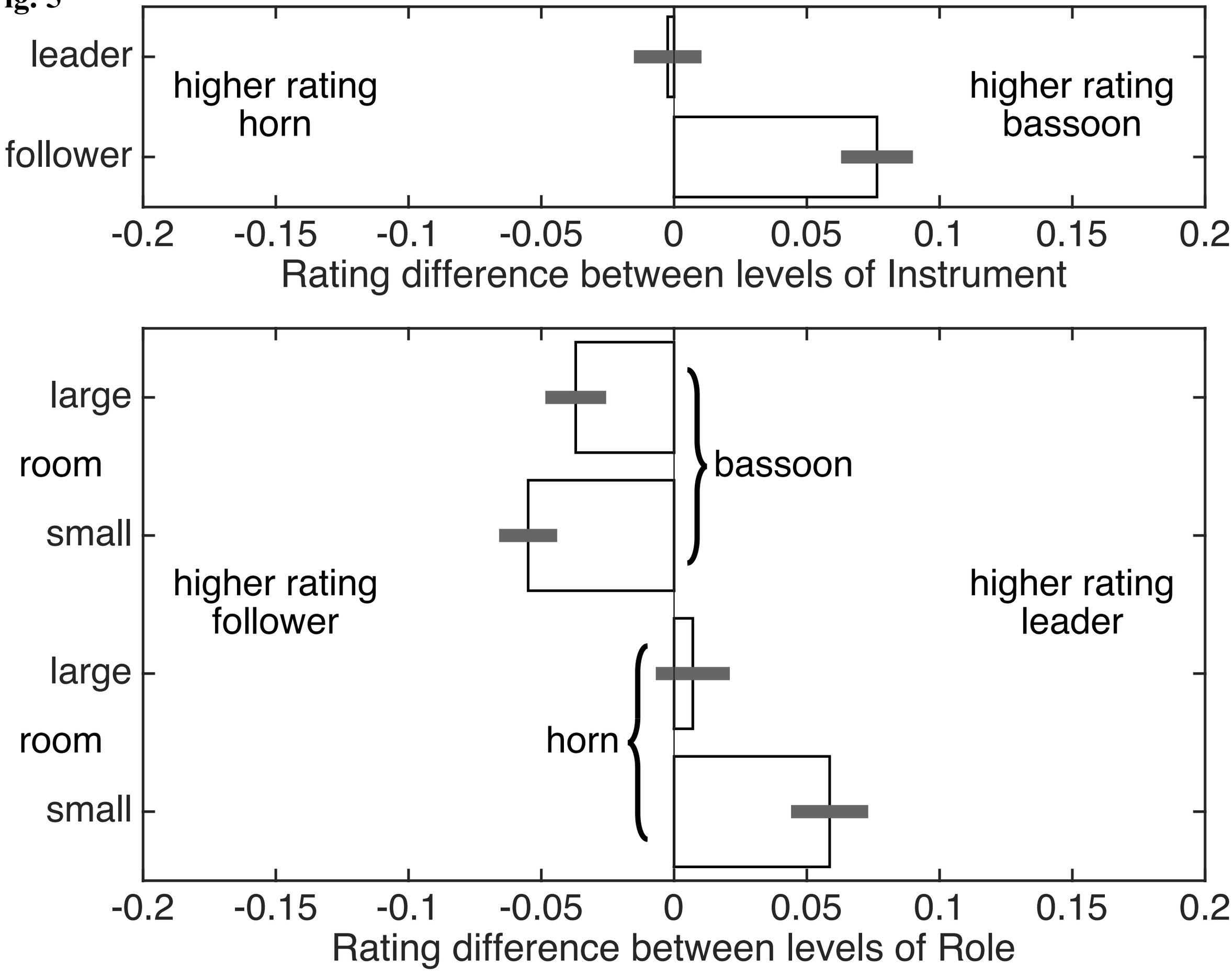

leader
follower
higher rating
horn
higher rating
bassoon
-0.2
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
0.2
Rating difference between levels of Instrument
large
room
small
bassoon
higher rating
follower
higher rating
leader
large
room
small
horn
-0.2
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
0.2
Rating difference between levels of Role

Fig. 6

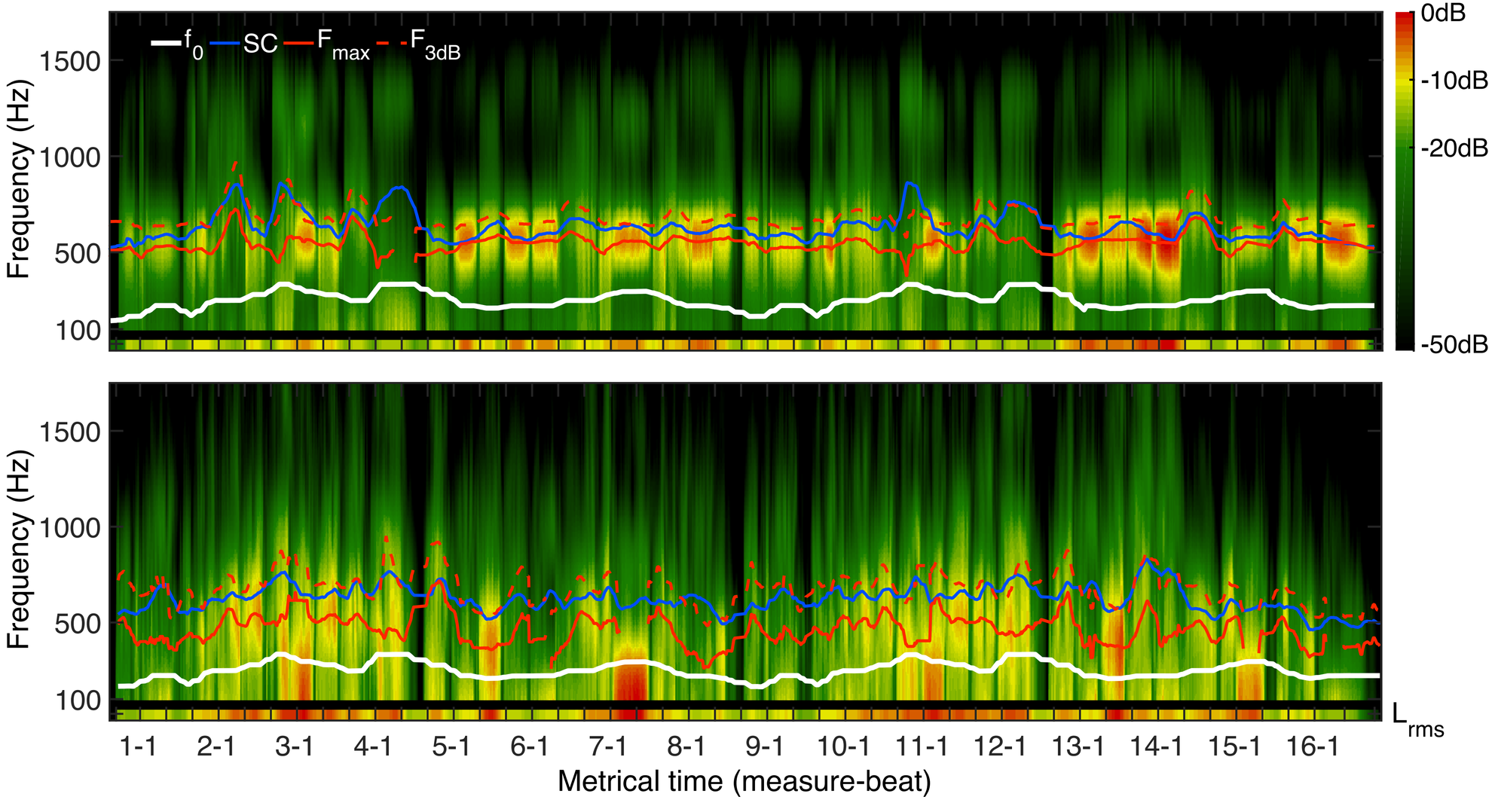

f0
SC
Fmax
F3dB
Frequency (Hz)
1500
1000
500
100
0dB
-10dB
-20dB
-50dB
Lrms
1-1
2-1
3-1
4-1
5-1
6-1
7-1
8-1
9-1
10-1
11-1
12-1
13-1
14-1
15-1
16-1
Metrical time (measure-beat)

Fig. 7

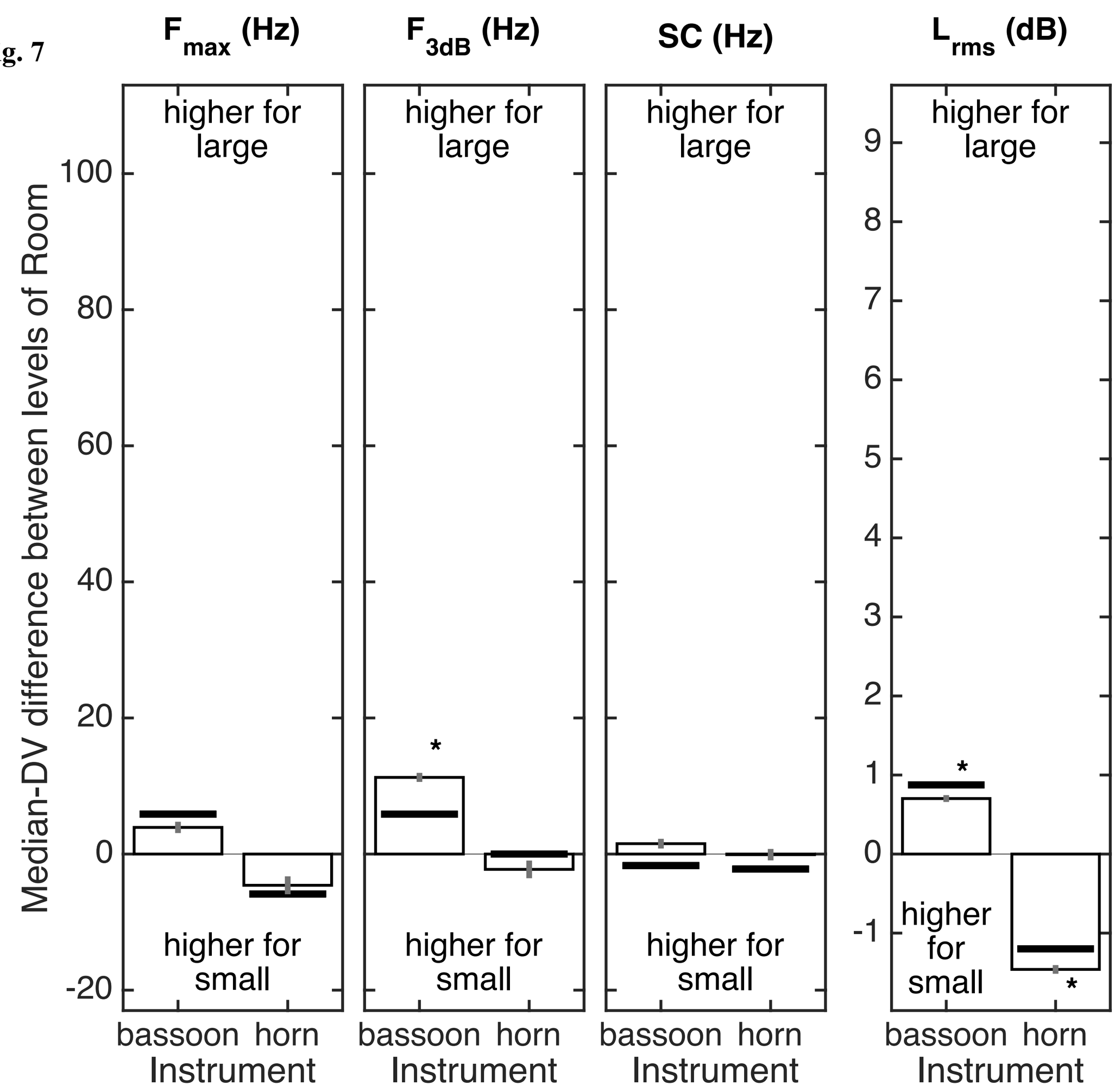

F max (Hz)
F 3dB (Hz)
SC (Hz)
L rms (dB)
higher for large
higher for small
Median-DV difference between levels of Room
100
80
60
40
20
0
-20
9
8
7
6
5
4
3
2
1
0
-1
*
bassoon horn
Instrument

Fig. 8

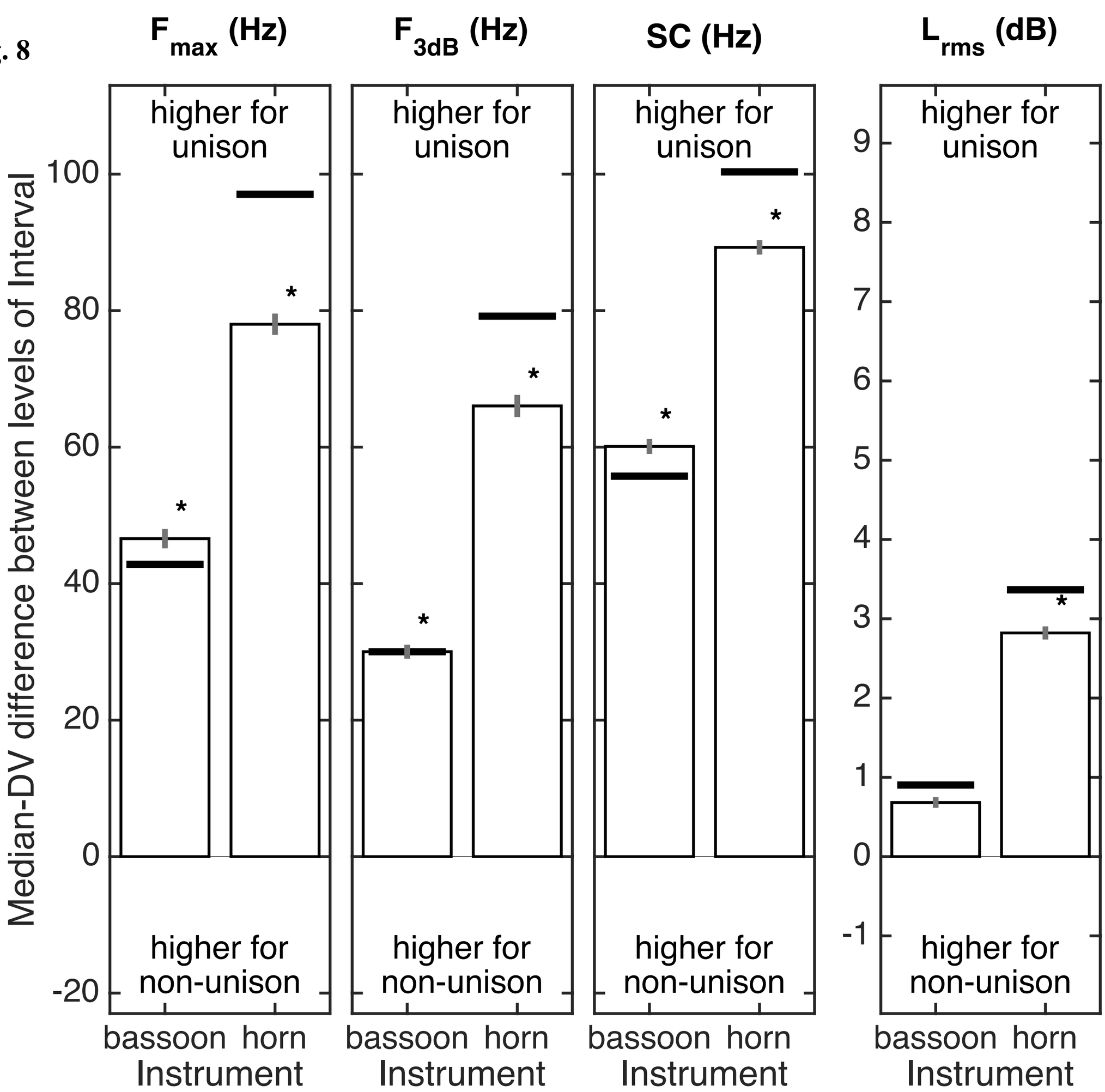

F max (Hz)
F 3dB (Hz)
SC (Hz)
L rms (dB)
higher for unison
higher for non-unison
Median-DV difference between levels of Interval
100
80
60
40
20
0
-20
9
8
7
6
5
4
3
2
1
0
-1
*
bassoon horn
Instrument

Fig. 9

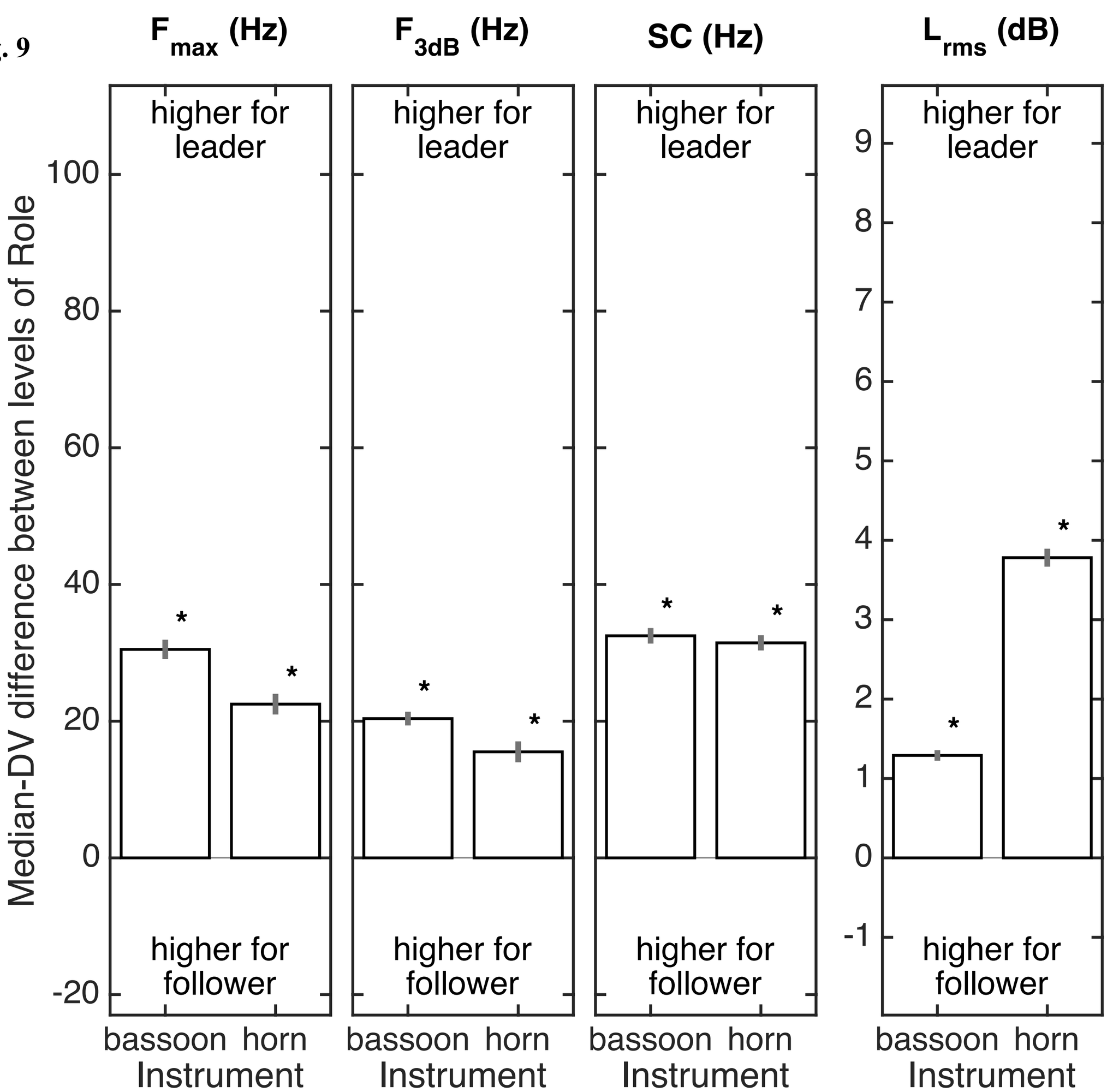

F max (Hz)
F 3dB (Hz)
SC (Hz)
L rms (dB)
higher for leader
higher for follower
Median-DV difference between levels of Role
100
80
60
40
20
0
-20
9
8
7
6
5
4
3
2
1
0
-1
*
bassoon horn
Instrument

Fig. 10

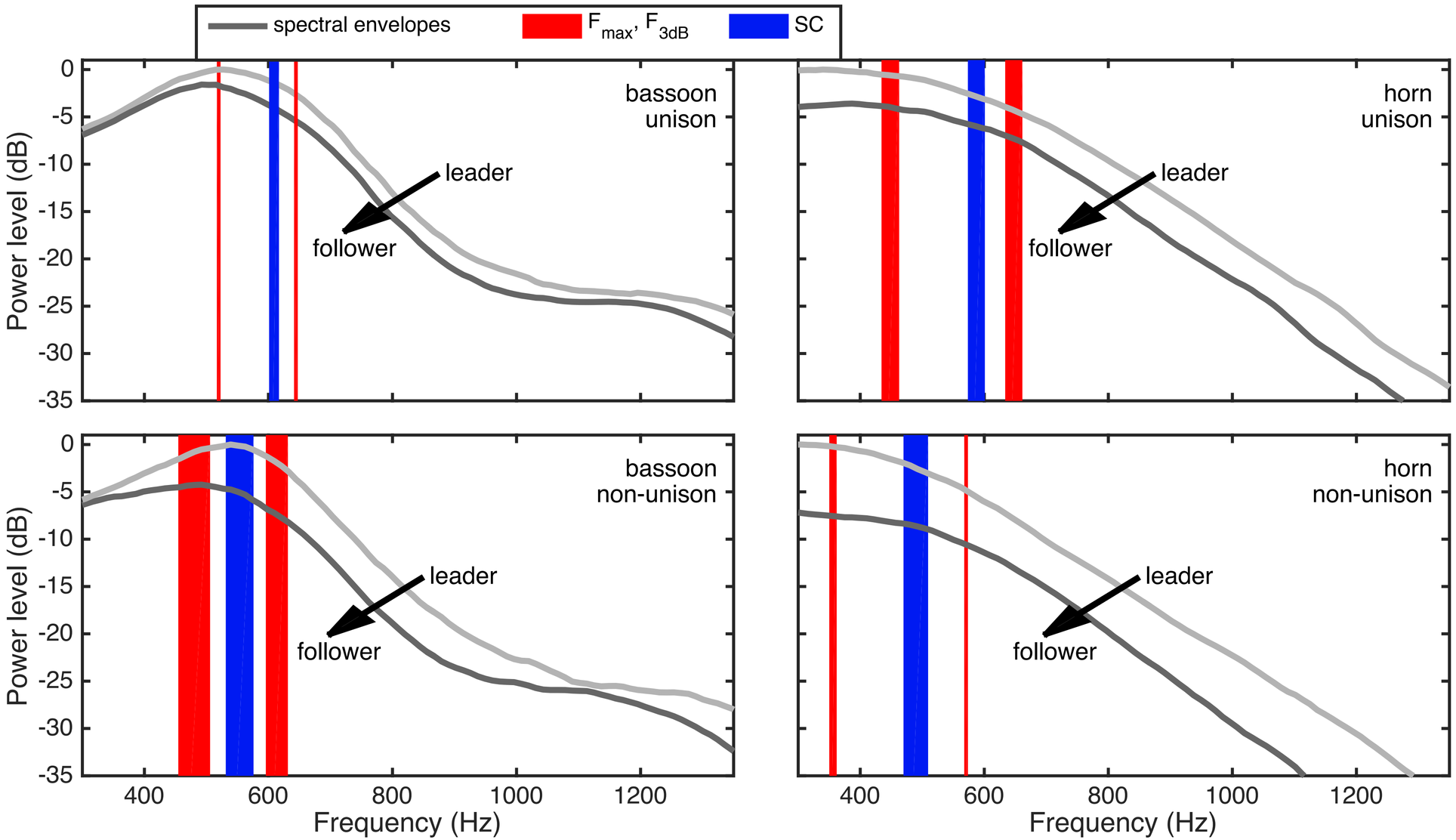

spectral envelopes
F max, F 3dB
SC
bassoon unison
horn unison
bassoon non-unison
horn non-unison
leader
follower
Power level (dB)
Frequency (Hz)
0
-5
-10
-15
-20
-25
-30
-35
400
600
800
1000
1200

Fig. 11

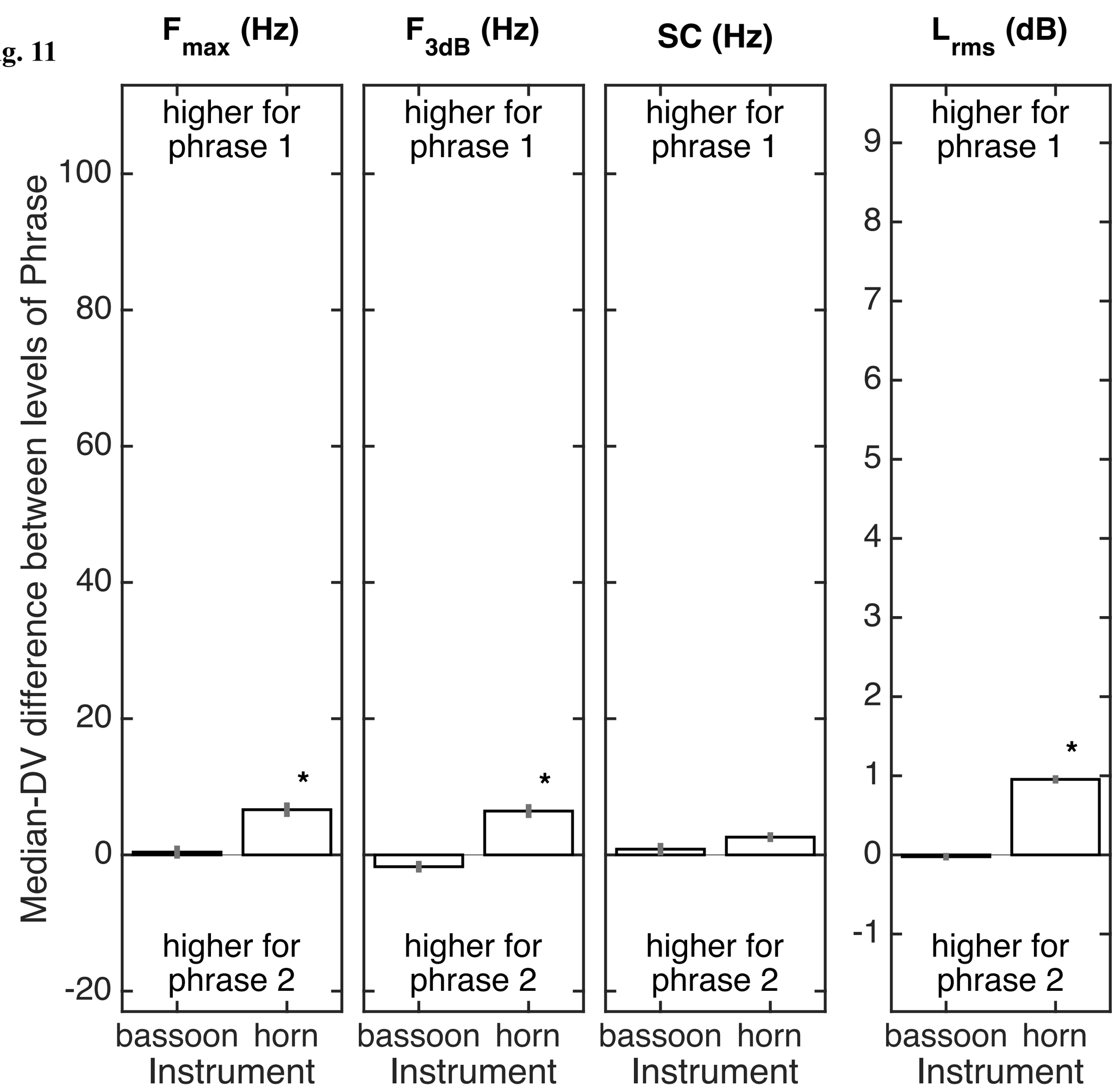
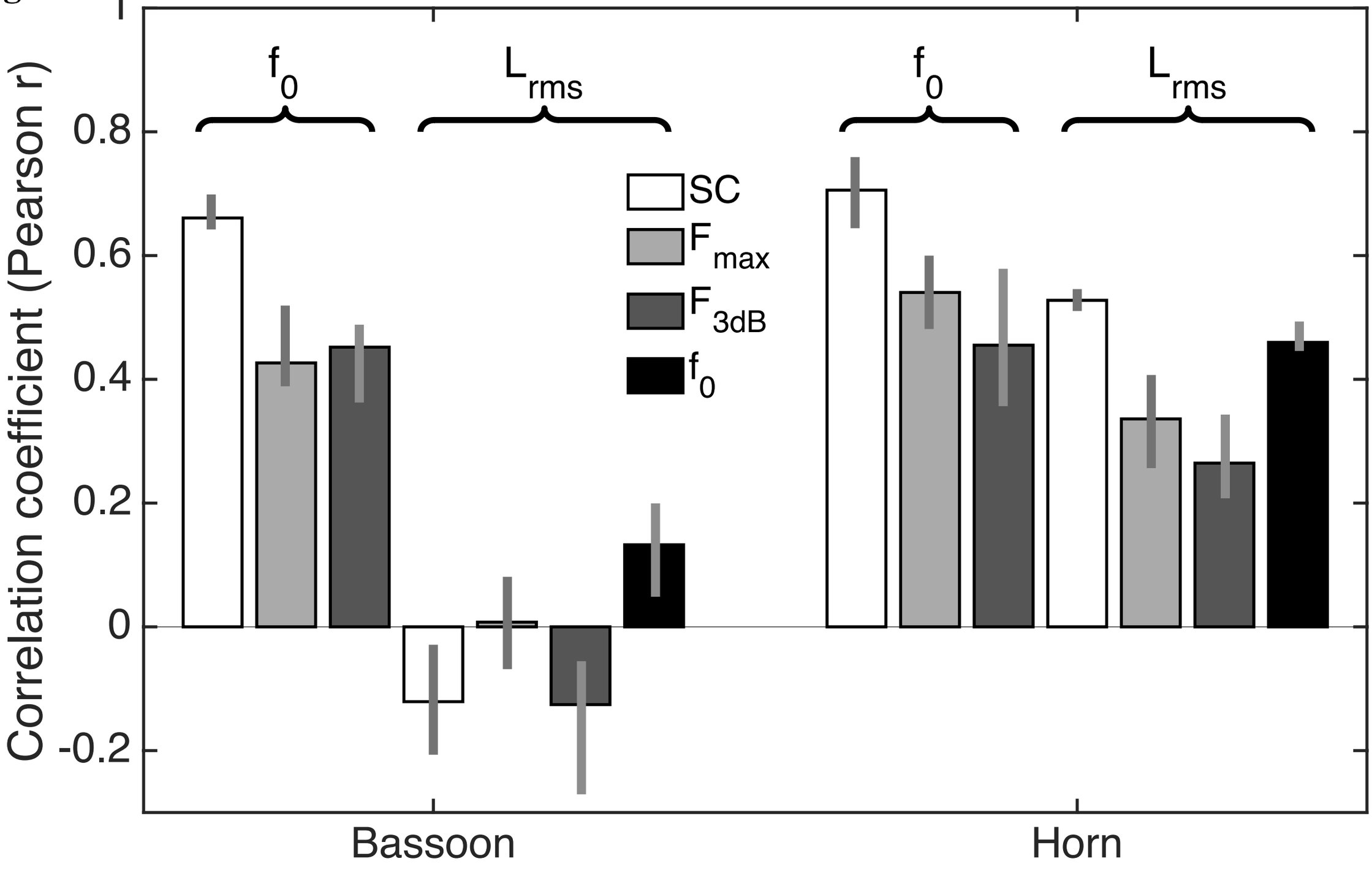

F max (Hz)
F 3dB (Hz)
SC (Hz)
L rms (dB)
higher for phrase 1
higher for phrase 2
Median-DV difference between levels of Phrase
100
80
60
40
20
0
-20
9
8
7
6
5
4
3
2
1
0
-1
*
bassoon horn
Instrument

Fig. 12


1
0.8
0.6
0.4
0.2
0
-0.2
Correlation coefficient (Pearson r)
f0
Lrms
f0
Lrms
SC
Fmax
F3dB
f0
Bassoon
Horn