**Supplementary Information**

Field Experiment Details

Field experiments were intentionally conducted within a short period during the dry season (March 2nd - 11th), to keep climatic variables as similar as possible (rainfall was 0 mm and mean temperature was 30.6˚C; range: 17.2 - 44.8˚C), since wood expands with heat and moisture resulting in variation in wood properties that can change acoustic properties [1]. Only six of the seven tree species observed for AST behaviour were used in the experiments because one (*Markhamia tomentosa*) was only clearly identified as different from another AST tree species (*Crossopteryx febrifuga*) at the end of the field season. For non-AST species, we chose seven species from the 28 remaining tree species clearly identified during reconnaissance surveys. These seven non-AST tree species were chosen because they were relatively abundant, have a DBH (diameter at breast height) greater than 35cm (the smallest mean DBH of an AST tree species; Table 1), as well as having no thorns or spikes protruding from their bark. We aimed to have all experiments completed before midday and randomised testing of AST and non-AST species within days when possible (Table S5).

Field Experiments- Standardized Throwing Gesture

To standardize throws, the microphone stand was always positioned at a horizontal distance of 1.4m from the base of the tree, AKK was positioned 1m from the centre of the microphone stand’s base and 1m horizontal distance from the base of the tree. AKK held the rock with her elbows resting against the inner sides of both knees whilst bent at the hips with legs straight. A target impact point for the tree was temporarily marked with flagging tape at 75cm above ground. All throws were underhand with the initial position of the rock being in-between AKK’s feet, hovering above the floor and parallel to her legs. The rock was then released once it was level with her knees. The dimensions and weight of all rocks thrown were measured beforehand and all presumed chimpanzee tools were handled with sterile gloves and put back in their original position. Only chimpanzee tools that had fresh impact signs at AST sites and were easily accessible were used to cause minimal disruption to sites. Two sound recording levels were used to strike a balance between a good signal to noise ratio and to avoid microphone clipping.

Acoustic Analyses- Details of Timbre Descriptors

The attack time characterizes the signal onset and corresponds to the time it takes for the signal to reach 90% of its maximum amplitude [1]. This sound descriptor correlates with the percussiveness of a sound and is one of the main auditory cues for the distinction between hard and soft impacts [2]. The spectral centroid is the centre of gravity of the modulus of the frequency spectrum. It is known to be strongly correlated with the perceived brightness of the sounds [3]. As opposed to the attack time, which is a temporal timbre descriptor, the spectral centroid is a spectral descriptor. Finally, the damping coefficient describes the global decay of the temporal signal, in other words the decrease of the sound energy as a function of time. The damping coefficient is strongly linked to the material properties of the impacted object and is an essential cue to distinguish one material category from another for the generation of sounds [2,4]. In our study, impact sounds result from the interaction between the thrown rock and the tree at a given excitation point. Attack time was multiplied by 1000 to ease interpretation of model estimates since absolute values produced by the algorithm were in milliseconds (Table S2). Similarly, the absolute damping coefficient was used since all values were negative due to the natural decreasing of the sounds (Table S4 and Figure 2).

Among the three acoustic descriptors used in this study, the attack time and the spectral centroid might be influenced by the hardness of the impact (i.e., throwing force) and the excitation point. However, since the thrower’s gesture was carefully controlled, the throwing force was comparably similar across throws. Note that despite our attempts to be standardized in throws, small variations in force could have occurred. As predicted by the modal analysis technique, substantial variations of impact force may influence the frequency range on which the resonances of the tree are excited, i.e., the harder the impact, the larger the frequency range. However, for small variations of force such as in our study, and considering that due to the relationship between the size of the trees and the size of the rocks, we are in the context of so-called linear vibrations. Only the global energy of the acoustic signals may be influenced, hereby the signal intensity, which was not considered as a descriptor in the analyses. The natural resonances of the tree excited by the impact (characterized by resonance frequencies and damping) are uncorrelated with the throwing force. Hence for a given type of rock and excitation point, we assume that the three timbre descriptors primarily characterize the intrinsic acoustic properties of the trees. For example, the internal friction of the wood species is linked to the way the sound decays (*Damping Coefficient*) as the sound energy is altered by both dispersion and dissipation phenomena which occur when acoustic waves propagate in the medium. The hardness of the tree bark at impact point is linked to the signal onset (*Attack Time*) and the modal response of the tree to the impact is linked to the center of gravity of the frequency spectrum (*Spectral Centroid*). In addition, the descriptor measurements are not influenced by recording level, meaning that no normalisation process was needed to analyse the impact sounds. The fact that attack time was not a significant descriptor for separating AST and non-AST trees but damping coefficient was, further support the hypothesis that any unconscious bias that may have resulted in variation in force by the experimenter did not obviously differ between AST and non-AST tree throws.

Statistical Analyses- Details of Linear Mixed Models

Models were fit with a Gaussian error distribution and default identity link function for LMMs since all response variables were continuous and assumed to follow a normal distribution [5,6]. These assumptions were verified by ensuring residuals were homogeneous and normally distributed via visual inspection of QQ-plots and plotting residuals against fitted values, indicating no violations. All LMMs were fit with the function ‘lmer’ of the package lme4 in R with the argument REML set to false to obtain maximum likelihood estimates [6,7]. All continuous fixed effects were z-transformed before running the models and all categorical predictors were dummy coded and centred. Random slopes for all fixed effects within the levels of the random effects were included when applicable (see Supplementary Information S2 for detailed R code of the Linear Mixed Models). We further checked for collinearity among fixed effects by running a linear model with no random effects and calculating the Variance Inflation Factors (VIFs) for all predictors [8] using the function ‘vif’ of the package car [9]. VIFs were between 1.01-1.31 demonstrating negligible collinearity. Model stability was verified by removing levels of random effects one at a time and ensuring model estimates did not vary strongly. Model significance was first assessed by conducting a full versus null model comparison using a likelihood ratio test with the function ‘anova’ set to a Chisq approximation [10]. If this showed significance (P<0.05) we determined the significance of individual test predictors using the ‘drop1’ function, again set to a Chisq approximation, to calculate likelihood ratio tests [5,11]. For significant categorical predictors, a post hoc test was conducted using the function ‘glht’ from the package multcomp with pairwise comparisons using a Tukey test [12].

To ensure our results were robust if only standardized rocks were used in the experiment, we further tested whether the removal of throws produced by rocks other than the standardized ones (i.e., S1, S2, S3; Table S1) changed any of our results. These LMMs could no longer include rock type because it is collinear with standardized rock ID, and standardized rock ID was fit as a fixed rather than random effect since three levels are insufficient for fitting random effects [5,6]. All other aspects of these LMMs remained the same, as described above. Despite the lower sample size (N=103 impact sounds) there was no change in significance for any of the full versus null model comparisons, nor any of the predictors, other than the effect of AST species on the absolute damping coefficient becoming P<0.01 rather than P<0.001 and the control variable of sound recording level no longer had a significant effect on damping measurements (Table S4).

References

1. Aramaki M, Baillères H, Brancheriau L, Kronland-Martinet R, Ystad S. 2007 Sound quality assessment of wood for xylophone bars. *The Journal of the Acoustical Society of America* **121**, 2407–2420. (doi:10.1121/1.2697154)

2. McAdams S, Winsberg S, Donnadieu S, De Soete G, Krimphoff J. 1995 Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychological Research* **58**, 177–192. (doi:10.1007/BF00419633)

3. Beauchamps JW. 1982 Synthesis by spectral amplitude and ‘brightness’ matching of analyzed musical intrument tones. *Journal of the Audio Engineering Society* **30**, 396–406.

4. Aramaki M, Besson M, Kronland-Martinet R, Ystad S. 2011 Controlling the perceived material in an impact sound synthesizer. *IEEE Transactions on Audio, Speech, and Language Processing* **19**, 301–314. (doi:10.1109/TASL.2010.2047755)

5. Dobson AJ, Barnett A. 2008 *An Introduction to Generalized Linear Models.* 3rd edition. Boca Raton, FL: Chapman and Hall/CRC.

6. Bates D, Mächler M, Bolker B, Walker S. 2015 Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software* **67**, 1–48. (doi:10.18637/jss.v067.i01)

7. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, Dai B, Grothendieck G, Green P. 2016 *lme4: Linear Mixed-Effects Models using ‘Eigen’ and S4*. See https://cran.r-project.org/web/packages/lme4/index.html.

8. Bowerman BL, O’Connell R. 2000 *Linear Statistical Models: An Applied Approach*. 2nd edition. Belmont, CA: Duxbury Press.

9. Fox J, Weisberg HS. 2001 *An R Companion to Applied Regression*. Thousand Oaks, CA: Sage Publications.

10. Forstmeier W, Schielzeth H. 2011 Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner’s curse. *Behav Ecol Sociobiol* **65**, 47–55. (doi:10.1007/s00265-010-1038-5)

11. Barr DJ, Levy R, Scheepers C, Tily HJ. 2013 Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* **68**, 255–278. (doi:10.1016/j.jml.2012.11.001)

12. Hothorn T, Bretz F, Westfall P, Heiberger RM, Schuetzenmeister A, Scheibe S. 2017 *multcomp: Simultaneous Inference in General Parametric Models*. See https://cran.r-project.org/web/packages/multcomp/multcomp.pdf.