**Supplemental Methods and Results**

Auditory perception: *Laurel* and *Yanny*, together at last

D. Pressnitzer, J. Graves, C. Chambers, V. de Gardelle, P. Egré

**Methods**

*Participants*. An online survey was set up with Qualtrics and advertised through social media and email lists. At the time of analysis, 289 volunteer participants had completed the survey. Before beginning the survey, all participants gave informed consent. The median total completion time of the survey was 7 minutes and 57 seconds.

*Stimuli*. The original soundclip was downloaded from a video posted on Twitter (<https://twitter.com/CloeCouture/status/996218489831473152>) and processed in MATLAB (The Mathworks, Inc., Natick, MA). Eleven versions of the sound were created by mixing highpass- and lowpass-filtered versions of the original soundclip. The highpass­ and lowpass filters were symmetrical, each a 4th-order Butterworth filter with a cutoff frequency of 2000 Hz. The eleven steps of the continuum were created by varying the amplitude ratio of the highpass- and lowpass-filtered components, from -60 to +60 dB in 12-dB increments, such that the 6th sound in the continuum was equivalent to the original sound.

*Procedure*. Each participant first answered an obligatory question about audio equipment, and was then prompted to answer optional questions about gender, age, native language, and level of musical training. During the main phase of the study, the eleven sounds of the continuum were each presented three times, in a randomized order, for a total of 33 trials. On each trial, participants were asked whether the sound they heard was closer to “Laurel” or “Yanny”. The instructions specified that words other than “Yanny” could be heard, such as “Yari” or “Yowee”, which were to be treated as “Yanny”. After the forced-choice response, participants were also asked to indicate the level of confidence in their response by moving a slider between the labels “Not Sure” and “Absolutely Sure”. The instructions specified that hearing both words simultaneously or neither should be reported as “Not sure”. At the end of the study, participants were asked to describe the sound in their own words, and to say whether they had heard the original soundclip prior to the experiment.

*Analysis.* Participants were separated into three groups on the basis of their overall response bias. The 134 participants who chose the same response on 90% or more of trials were classified as either “Laurelists” or “Yannists” (N = 93 Laurel, N = 41 Yanny). The remaining 155 participants were analyzed separately as the “Intermediate” group. Two 2-way mixed ANOVAs were then conducted, with filter condition as a within-subjects factor and group as a between-subjects factor, one on the Laurel/Yanny response and one on confidence.

To look for correlations with various items of the survey, we chose to consider responses at the high/low filter ratio of +12 dB, the most ambiguous point we tested on the continuum when considering all responses. We did so for both “Laurel/Yanny” forced choice responses and the confidence ratings.

In order to characterize the variability in the point of transition between the “Laurel” and “Yanny” responses, a sigmoid curve was fitted to the responses of each participant in the Intermediate group, and we plotted histograms of the inflection point of this curve, and confidence at that inflection point for each participant (Fig. 1D and 1E). It should be noted that each participant completed only 3 trials at each high/low ratio, which is less than ideal for psychometric curve estimation. However, even with this limited number of trials at the individual level, visual inspection of the data suggested that the peak and distribution of inflection points were interpretable.

To assess the effect of context from previous trials, we performed two binary logistic regressions on the Laurel/Yanny response data from each individual in the Intermediate group. The first model predicted the response on the current trial using the current stimulus and the response from the previous trial as predictors. The second model predicted the response on the current trial using the current stimulus and the stimulus from the previous trial as predictors.

**Results**

*High/low ratio and overall Laurel/Yanny bias.* A 2-way ANOVA on the Laurel/Yanny responses found an effect of group [F(2,286) = 1,167.45, p < 0.001, η² = 0.89], an effect of filter condition [F(10,2860) = 201.41, p < 0.001, η² = 0.41], and an interaction [F(20,2860) = 235.35, p < 0.001, η² = 0.62]. A 2-way ANOVA on confidence found the same three effects (group: [F(2,286) = 5.48 p = 0.0046, η² = 0.037], filter condition: [F(10,286) = 22.96, p < 0.001, η² = 0.074], interaction: [F(20,2860) = 10.05, p < 0.001, η² = 0.066]).

*Confidence in the dominant vs. non-dominant responses.* To confirm that participants used the confidence scale in a sensible manner, we compared their confidence in the dominant response vs. the non-dominant response. We expected that confidence would be higher in the dominant response, a classic finding in psychophysical studies where the dominant response is typically the correct response. Note that because many participants used only one response throughout the experiment, this pattern could only be tested on a subset of participants who used both responses (10 Yannists out of 41, 23 Laurelists out of 93, and all 155 participants in the Intermediate group). Within this subset, we found that Laurelists were indeed more confident in their “Laurel” responses than in their “Yanny” responses [t(22)= 3.975, p=0.00064]. By contrast, Yannists were more confident in their “Yanny” responses [t(9)=1.330, p=0.21610] although this was not significant, probably due to the low number of participants in this subgroup. Between these two groups, the confidence gap for the “Laurel” vs. “Yanny” responses therefore had opposite directions, and significantly differed [t(31)=3.253, p=0.00276]. Moreover, the relation between response dominance and confidence was also confirmed when focusing on the intermediate group, where we found that the proportion of “Laurel” responses significantly predicted the confidence gap between “Laurel” and “Yanny” responses, across individuals [β=62.25, SE=11.01, p<0.001].

*Audio equipment.* A one-way ANOVA on p(Laurel) at +12 dB, including only the between-subjects factor of reported audio equipment (with four categories: (“over-ear headphones”, “in-ear earbuds”, “loudspeaker (mobile phone)” and “loudspeaker (laptop computer)”), found a significant effect [F(3,268) = 4.17, p = 0.0066, η² = 0.045]. Pairwise comparisons with Bonferroni correction revealed that both over-ear headphones and in-ear earbuds produced more Laurel responses at +12 dB than laptop computer loudspeakers, with no other comparisons reaching significance. A similar ANOVA on confidence at +12 dB also found a significant effect [F(3,268) = 3.51, p = 0.016, η² = 0.038]. Pairwise comparisons revealed that mobile phone loudspeakers produced higher confidence at +12 dB than in-ear earbuds, with no other comparisons reaching significance.

*Learning and experience.* We measured two potential effects of learning and experience: familiarity with the specific sound used in the experiment, and musical experience. Participants who had heard the sound before our experiment were more likely to respond “Laurel” at +12 dB [t(287) = 2.72, p = 0.0069, Cohen’s d = 0.36], and were also more confident at +12 dB [t(287) = 3.87, p < 0.001, Cohen’s d = 0.48]. Participants who reported prior musical training were more likely to respond “Laurel” at +12 dB [t(284) = 2.05, p = 0.041, Cohen’s d = 0.24], but not more or less confident at +12 dB than those who did not [t(284) = -0.36, p = 0.72, Cohen’s d = 0.043]. Using Spearman’s rank correlation to account for non-normally distributed data, a significant correlation was observed between reported years of musical training and p(Laurel) at +12 dB [Spearman’s ρ= 0.13, p = 0.03], but no correlation was observed between years of musical training and confidence at +12 dB [Spearman’s ρ= 0.027, p = 0.65].

*Demographic factors.* Out of 289 participants, 286 reported their age, and 286 reported their gender as either male or female. No correlations were found with participants’ reported age, neither for p(Laurel) at +12 dB [Spearman’s ρ= 0.036, p = 0.54] nor for confidence at +12 dB [Spearman’s ρ= 0.092, p = 0.12]. We observed a significant effect of gender on p(Laurel) at + 12 dB, such that men were more likely than women to respond “Laurel” [t(284) = 2.46, p = 0.015, Cohen’s d = 0.29], as well as an effect of gender on confidence at +12 dB, such that women were more confident than men [t(284) = 2.26, p = 0.025, Cohen’s d = 0.27]. Participants were split into three groups by native language: native English speakers (N=118), native French speakers (N=99), and other (N=72). A one-way ANOVA on p(Laurel) at +12 dB found a significant effect of native language [F(2,286) = 4.23, p = 0.015, η² = 0.029]. Pairwise comparisons revealed that native English speakers were significantly more likely to report “Laurel” than native French speakers, with no other comparisons reaching significance. A similar ANOVA on confidence at +12 dB found a marginally significant effect [F(2,286) = 3.14, p = 0.045, η² = 0.021], but no pairwise comparisons were significant.

*Context from previous trials.* The first logistic regression model found a significant effect of current stimulus [z = 36.57, p < 0.001], as well as of the previous response [z = 10.01, p < 0.001], confirming that the response on each trial was influenced by both the stimulus on that trial and the response on the previous trial. The second logistic regression model found an effect of current stimulus [z = 37.05, p < 0.001], and a small but significant negative effect of the previous stimulus [z = -2.14, p = 0.03]. The small negative effect of the previous stimulus could be interpreted as an effect of contrast. The small size of this effect is perhaps due to the presentation mode, where the number of repetitions of the sound per trial and the timing were not controlled. Another independent online study using stimuli specifically designed to reveal spectral contrast effects claimed such an effect of acoustic context (https://www.psychologytoday.com/us/blog/chitchat-about-chatter/201805/the-psychology-laurel-and-yanny?amp)