Putting Laurel and Yanny in context

Hans Rutger Bosker

Citation: The Journal of the Acoustical Society of America **144**, EL503 (2018); doi: 10.1121/1.5070144
View online: https://doi.org/10.1121/1.5070144
View Table of Contents: http://asa.scitation.org/toc/jas/144/6
Published by the Acoustical Society of America

Hans Rutger Bosker: JASA Express Letters https://doi.org/10.1121/1.5070144 Published Online 5 December 2018

Putting Laurel and Yanny in context

Hans Rutger Bosker[a)]
Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH, Nijmegen, The Netherlands
HansRutger.Bosker@mpi.nl

Abstract: Recently, the world's attention was caught by an audio clip that was perceived as "Laurel" or "Yanny." Opinions were sharply split: many could not believe others heard something different from their perception. However, a crowd-source experiment with >500 participants shows that it is possible to make people hear Laurel, where they previously heard Yanny, by manipulating preceding acoustic context. This study is not only the first to reveal within-listener variation in Laurel/Yanny percepts, but also to demonstrate contrast effects for global spectral information in larger frequency regions. Thus, it highlights the intricacies of human perception underlying these social media phenomena.

[DDO]
Date Received: June 18, 2018 **Date Accepted:** October 16, 2018

1. Introduction

In May 2018, social media exploded after the surfacing of an audio clip that some perceived as "Laurel," but others as "Yanny." The clear divide between #Yannists and #Laurelites was reminiscent of #TheDress, a photo going viral in 2015 of a white and gold dress, or was it black and blue (Brainard and Hurlbert, 2015)? Although some referred to the auditory Laurel/Yanny phenomenon as "black magic," critical observers noticed unusually strong higher frequencies (confirmed by the acoustic analysis below) which would resemble the acoustic signature of Yanny. This could potentially explain the variation between listeners, for instance, due to electronic devices varying in how they represent the higher frequencies, or due to diminished peripheral hearing with age, typically with the largest decrements in the higher frequencies (presbycusis; Gates and Mills, 2005). Still, perception is never an objective registration of sensory information: it draws upon information from prior experience and context. To demonstrate this, the present study shows that it is possible to make people hear Yanny, where they previously heard Laurel, by manipulating the frequency content in the surrounding acoustic context.

Contextual contrast enhancement is a fundamental processing principle in many species, allowing perceivers to navigate their highly variable environment, relying on relative, rather than absolute, coding strategies. Examples from speech processing are temporal (Bosker, 2017; Reinisch and Sjerps, 2013) and spectral contrast effects (Ladefoged and Broadbent, 1957; Sjerps *et al.*, 2011), whereby a preceding acoustic context influences following target categorization. For instance, lowering the first formant in a precursor leads to the perception of a higher first formant in the following target (Ladefoged and Broadbent, 1957). Studies on spectral contrast have typically targeted contrastive perception of specific formants or formant transitions (Lotto and Kluender, 1998). As such, it remains unknown whether spectral contrast also applies to the perception of more global spectral information in much larger frequency regions.

Therefore, the present online crowd-source experiment presented listeners with clips from a 7-step phonetic continuum (from 1, most Laurel-like, to 7, most Yanny-like), modulating the intensity of lower vs higher frequencies. Moreover, each clip was preceded by a lead-in sentence (a precursor consisting of a 7-digit telephone number: "496-0356"; cf. Bosker and Ghitza, 2018) that was either low-pass filtered (attenuating frequencies >1000 Hz), as if overhearing someone in the room next door; or high-pass filtered (attenuating frequencies <1000 Hz), as if hearing someone over a phone. It was predicted that artificially boosting the higher frequencies in the phonetic continuum would bias perception toward Yanny. Furthermore, following the principles of spectral contrast, attenuating higher frequencies in the precursor would make the higher frequencies in the following clip stand out more, leading to more Yanny responses.

[a)]Also at: Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands.

2. Methods

2.1 Acoustic analysis

Figure 1(b) shows the spectrogram of the original audio clip that went viral on social media. The clip has been said to have been recorded from a set of speakers, playing the noun "laurel" pronounced by a male native speaker of English from a vocabulary website: www.vocabulary.com (as explained by Wired; Matsakis, 2018). Therefore, Fig. 1 compares the original audio clip [Fig. 1(b)] to the presumed source sound [Fig. 1(a)], and a "simulated" Yanny recording in the same voice [Fig. 1(c)]. The Yanny recording was created by splicing together the first two sounds of "yank" (although the /æ/ in "yank" vs "yanny" may differ in phonetic realization depending on regional dialect, they share the same phoneme) and the last two sounds of "uncanny" in the same voice as the original audio clip. Comparison of the spectrograms reveals that the higher frequencies (>1000 Hz) are relatively enhanced in the original clip, and in particular the pronounced resonance (frequencies with higher amplitude, shown in red shading) dropping from 3 kHz to below 2 kHz in the first 400 ms. Because the exact recording conditions are unknown, it is unclear what caused this (natural/artificial) enhancement.

It could be that this pronounced resonance is interpreted by some as the third formant, perceiving Laurel (e.g., male /ɔ/ typically has an $F3$ around 2410 Hz; Peterson and Barney, 1952), but by others as the second formant, perceiving Yanny (e.g., male /i/ typically has an $F2$ around 2290 Hz; Peterson and Barney, 1952). Also, given the close proximity (in frequency space) of the lower formants (around 500 and 900 Hz, respectively), confusion may arise in their perception (especially with relatively loud higher frequencies): either as one formant with a large bandwidth, or two with smaller bandwidths.

Comparison of the original clip [Fig. 1(b)] to a simulated Yanny recording [Fig. 1(c)] reveals some similarity particularly in the high frequency content of the two sounds. In fact, the intensity of frequencies >1000 Hz shows more similarity to Yanny than to Laurel. A cross-correlation of the power spectra of low-pass filtered (cutoff frequency = 1000 Hz, using a Hann window with a roll-off width of 100 Hz as implemented in Praat) versions of the sounds (all with a sampling frequency of 44.1 kHz) shows a stronger maximal correlation (with a much smaller frequency shift) between the lower frequencies of the original clip and those of Laurel ($r = 0.822$; frequency shift = 0.008 kHz; $p < 0.001$), compared to Yanny ($r = 0.691$; frequency shift = 21 kHz; $p < 0.001$). Conversely, the power spectra of high-pass filtered (cutoff frequency = 1000 Hz, using a Hann window with a roll-off width of 100 Hz as implemented in Praat) versions of the sounds show a slightly stronger maximal correlation (with a considerably smaller frequency shift) between the higher frequencies of the original clip and those of Yanny ($r = 0.607$; frequency shift = −6 kHz; $p < 0.001$), compared to Laurel ($r = 0.606$; frequency shift = −17 kHz; $p < 0.001$).

At the request of an anonymous reviewer, we attempted to recreate a similar phenomenon with new stimuli in order to showcase that cues to different words in different regions of frequency space can result in perceptual ambiguity. A male native

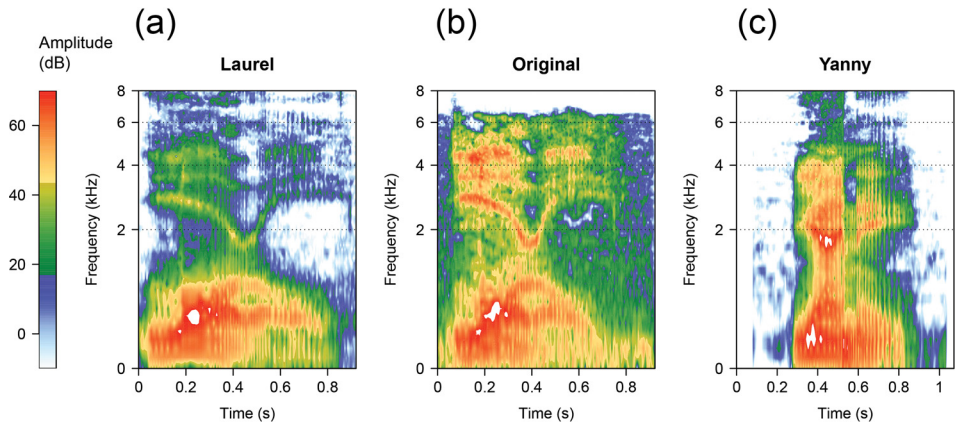

Fig. 1. (Color online) Comparison of the original clip to its source sound Laurel and a simulated Yanny. (a) Spectrogram of the source sound: a male native speaker of English pronouncing the noun laurel on www.vocabulary.com. (b) Spectrogram of the original clip that went viral on social media, obtained from Twitter. (c) Spectrogram of a simulated Yanny, in the same voice as the original clip, created by splicing together the first two sounds of yank and the last two sounds of uncanny.

speaker of English was recorded producing the names Harry and Meghan. The lower frequencies of Harry and the higher frequencies of Meghan were combined by low-pass filtering Harry and high-pass filtering Meghan (using Hann windows with a roll-off width of 100 Hz) at three different cutoff frequencies. By shifting the cutoff frequency, perception is shifted toward one or the other of the two names, with Audio S1 (cutoff frequency 500 Hz, i.e., only few lower frequencies from Harry) sounding mostly Meghan-like; Audio S2 (cutoff frequency 2000 Hz, i.e., cues to both names in the lower vs higher frequencies) sounding ambiguous between Harry and Meghan; and Audio S3 (cutoff frequency 3200 Hz; most cues, particularly in lower frequencies, to Harry) sounding most Harry-like (see the supplementary material[1]).

2.2 Participants

Participants ($N = 532$) were recruited through a blog post on Psychology Today, through the website of the Max Planck Institute, and through personal communication. Participants gave informed consent as approved by the Ethics Committee of the Social Sciences department of Radboud University (project code: ECSW2014-1003-196). Participant data were collected anonymously; as a result, the participants' gender, age, native language, location, sound device, etc., are unknown.

2.3 Materials and design

A phonetic continuum was created from the original clip (obtained from Twitter; sampling frequency = 44 100 Hz). The original clip was placed at step 4 (0 dB emphasis/attenuation). This original clip was filtered by 10 bandpass filters (with center frequencies: 31.5, 63, 125, 250, 500, 1000, 2000, 4000, 8000, 16 000 Hz; using a Hann window with a roll-off width of 20, 20, 40, 80, 100, 100, 100, 100, 100, 100 Hz, respectively). The output of the filters was manipulated in intensity: one step up the continuum represents a +6 dB emphasis on frequencies >1000 Hz and −6 dB attenuation on frequencies <1000 Hz [Fig. 2(a); implemented using Praat; Boersma and Weenink, 2016]. After combining the manipulated frequency bands back together, resulting tokens were matched in intensity to the original clip. Also, another male native speaker of English (i.e., a different voice) was recorded producing the telephone number 496–0356, which was subsequently low-pass filtered (1000 Hz cutoff, using a Hann window, roll-off width = 100 Hz) and high-pass filtered [1000 Hz cutoff, using a Hann window, roll-off width = 100 Hz; Fig. 2(b)], after which overall intensity was matched. All continuum steps were combined with either precursor, resulting in 14 unique stimuli. Each of these 14 unique stimuli was presented 5 times in random order to participants using the online tool PsyToolkit (Stoet, 2010). Although no control could be exerted over the type of electronic device (mobile phones, tablets, laptops), audio equipment (headphones,

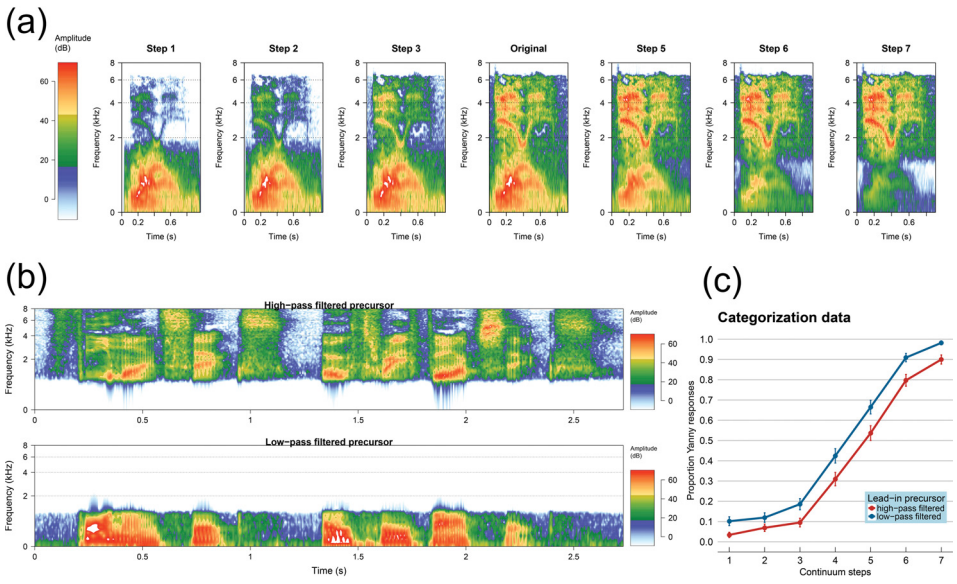

Fig. 2. (Color online) Materials and results from crowd-source experiment. (a) Spectrograms of the 7-step phonetic continuum. (b) High- and low-pass filtered versions of the precursor telephone number 496–0356. (c) Average categorization data (in proportion Yanny responses) for each step on the phonetic continuum, separately for the two precursor conditions (low-pass filtered vs high-pass filtered speech). Error bars enclose 1.96 × SE on either side; 95% confidence intervals.

speakers), or what browser participants used, participants were explicitly instructed to use headphones. Their task was to indicate what the final word was: Laurel or Yanny (self-paced). Audio clips from the phonetic continuum (Audio S4-S10) and the two precursors (Audio S11-S12) have been made available as online supplementary material.[1]

3. Results

Data and analysis scripts are available for download from osf.io/63wdh. Requests for further resources should be directed to and will be fulfilled by the author.

First, participants that only completed fewer than 5 trials were excluded from analyses (<1% data loss), after which the average number of trials completed across participants was 37. Of the remaining 384 participants, 48 reported almost exclusively Laurel (<0.1 proportion Yanny responses) and 60 reported almost exclusively Yanny (>0.9 proportion Yanny responses), highlighting some of the perceptual stability (i.e., hearing the same word in over 90% of the cases, independent from the continuum and context manipulations) reported on social media.

Categorization data from the other 276 participants (that did show variation in categorization), calculated as the proportion of Yanny responses, are presented in Fig. 2(c). A Generalized Linear Mixed Model with a logistic linking function tested the binomial dependent variable (participants' categorization of the final word as either Yanny, coded as 1, or as Laurel, coded 0) for fixed effects of Continuum Step (continuous predictor; centered and scaled around the mean), Precursor Condition (categorical predictor; deviation coding, with high-pass filtered speech coded as −0.5 and low-pass filtered speech as +0.5), and their interaction. Random intercepts for participants were also included.

This model revealed significant effects of Continuum Step [$\beta = 3.036$, standard error (SE) = 0.063, $z = 48.490$, $p < 0.001$; higher proportion Yanny responses for higher steps on the continuum] and Precursor Condition ($\beta = 1.050$, SE = 0.071, $z = 14.740$, $p < 0.001$; higher proportion Yanny responses after low-pass filtered speech), but no interaction ($p = 0.355$).

4. Discussion

The acoustic analysis suggests that the higher frequencies contain more phonetic cues to Yanny, whereas the lower frequencies contain more cues to Laurel. The categorization data show that the original clip (Step 4) was indeed ambiguous between Laurel and Yanny, with a slight bias in the present participant sample to report Laurel (in over 60% of the cases) over Yanny (over 30% of the time). The data also show that emphasizing the higher frequencies, while at the same time attenuating the lower frequencies, biases perception toward Yanny. Attenuating the lower frequencies may possibly have led participants to interpret the lower resonances as one first formant with a large bandwidth. Moreover, louder higher frequencies would make the clip resemble the acoustic signature of Yanny to a greater degree.

Note, however, that a sizable minority was insensitive to the experimental manipulations: 108 participants showed perceptual stability in their categorization data, almost exclusively reporting one of the two response options. This variation between participants may be accounted for (in part) by the lack of control over experimental conditions in online testing. If participants used low quality sound devices that poorly represented the higher frequencies, this would negatively affect their sensitivity to the experimental manipulations. Similarly, individuals suffering from (mild to severe) hearing loss (e.g., presbycusis), which typically affects the higher frequencies to a greater extent, would also be expected to be less sensitive to the experimental manipulations. This perceptual stability in a subset of the participants is likely also why the Laurel/Yanny phenomenon caught so much attention: many posts on social media commented on disbelief about others hearing something different from one's own perception. Future studies may disentangle exactly how stimulus features and listener characteristics interact in the perception of Laurel/Yanny.

At the same time, the present data demonstrated, for the first time, within-listener variation in Laurel/Yanny perception by means of an acoustic context manipulation. Listeners were more likely to categorize the same clip as Yanny after a low-pass filtered precursor, but as Laurel after a high-pass filtered precursor. Note that the precursor was spoken in a different voice from the target audio clip; still, the availability of global spectral information in large frequency regions in the precursor influenced the perception of another talker. This suggests that the context effect observed is (at least in part) driven by general acoustic processes (Lotto and Kluender, 1998; Sjerps *et al.*, 2011; Stilp *et al.*, 2010), with the surrounding acoustic environment

influencing how any one listener perceives any one Laurel/Yanny stimulus at a given time.

The precursor effect observed in the present data is in line with literature on acoustic context effects in audition, and spectral contrast effects in particular (Assgari and Stilp, 2015; Bosker *et al.*, 2017; Feng and Oxenham, 2018; Holt *et al.*, 2000; Sjerps and Reinisch, 2015). For instance, a higher second formant in a preceding context biases listeners to perceive a lower second formant in a subsequent target (Bosker *et al.*, 2017). Spectral contrast effects have been explained in terms of neuronal adaptation: the listening brain is known to respond to spectral regularities in the long-term statistics of acoustic signals by depressing neuronal responses to regularity and, conversely, enhancing responses to auditory novelty (Holt, 2006; Huang and Holt, 2012). Adopting this adaptive coding framework, the present findings suggest that neural responses to higher frequencies in the target words were enhanced as a result of exposure to the low-pass filtered precursor (and vice versa for the low-pass filtered precursor), heightening listeners' sensitivity to the phonetic cues in the targets' higher frequencies, biasing perception toward Yanny.

Interestingly, earlier studies on spectral normalization have typically targeted contrastive perception of specific formants (Bosker *et al.*, 2017), formant transitions (Lotto and Kluender, 1998), or spectral tilt (Alexander and Kluender, 2010; Kiefte and Kluender, 2008). To our knowledge, this is the first demonstration that the presence or absence of global spectral information in larger frequency regions contrastively influences the perception of subsequent spectral information in that region. At the same time, the present study is the first to demonstrate within-listener variation in Laurel/Yanny percepts, contrary to the widely reported perceptual stability of Laurel/Yanny perception on social media. Thus, it highlights the subjective and context-dependent nature of perception and shows that #Laurelgate can be instrumental in helping us understand the intricacies of human perception.

Acknowledgments

The author was supported by a Gravitation grant from the Dutch Government to the Language in Interaction Consortium. The author would like to thank Antje Meyer and Matthias Sjerps for useful comments on an earlier draft, Joe Rodd for help with data visualization, and Merel Maslowski for pointing the author to the original Laurel/Yanny post on Twitter.

References and links

[1]See supplementary material at https://doi.org/10.1121/1.5070144 for example sounds: Audio S1-3 demonstrate the Harry/Meghan items; Audio S4-10 demonstrate the Laurel/Yanny phonetic continuum; Audio S11-S12 demonstrate the two precursors used in the perception experiment.

Alexander, J. M., and Kluender, K. R. (**2010**). "Temporal properties of perceptual calibration to local and broad spectral characteristics of a listening context," J. Acoust. Soc. Am. **128**(6), 3597–3613.

Assgari, A. A., and Stilp, C. E. (**2015**). "Talker information influences spectral contrast effects in speech categorization," J. Acoust. Soc. Am. **138**, 3023–3032.

Boersma, P., and Weenink, D. (**2016**). Praat: Doing phonetics by computer [computer program].

Bosker, H. R. (**2017**). "Accounting for rate-dependent category boundary shifts in speech perception," Attn., Percept. Psychophys. **79**, 333–343.

Bosker, H. R., and Ghitza, O. (**2018**). "Entrained theta oscillations guide perception of subsequent speech: Behavioural evidence from rate normalization," Lang., Cognition Neurosci. **33**(8), 955–967.

Bosker, H. R., Reinisch, E., and Sjerps, M. J. (**2017**). "Cognitive load makes speech sound fast but does not modulate acoustic context effects," J. Memory Lang. **94**, 166–176.

Brainard, D. H., and Hurlbert, A. C. (**2015**). "Colour vision: Understanding #TheDress," Current Biol. **25**(13), R551–R554.

Feng, L., and Oxenham, A. J. (**2018**). "Spectral contrast effects produced by competing speech contexts," J. Exp. Psychol. **44**(9), 1447–1457.

Gates, G. A., and Mills, J. H. (**2005**). "Presbycusis," Lancet **366**(9491), 1111–1120.

Holt, L. L. (**2006**). "The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization," J. Acoust. Soc. Am. **120**, 2801–2817.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722.

Huang, J., and Holt, L. L. (**2012**). "Listening for the norm: Adaptive coding in speech categorization," Frontiers Psychol. **3**, 1–6.

Kiefte, M., and Kluender, K. R. (**2008**). "Absorption of reliable spectral characteristics in auditory perception," J. Acoust. Soc. Am. **123**(1), 366–376.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Lotto, A. J., and Kluender, K. R. (**1998**). "General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification," Percept. Psychophys. **60**, 602–619.

Matsakis, L. (**2018**). *The True History of "Yanny" and "Laurel*," https://www.wired.com/story/yanny-and-laurel-true-history/ (Last viewed August 2, 2018).

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Reinisch, E., and Sjerps, M. J. (**2013**). "The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context," J. Phonetics **41**, 101–116.

Sjerps, M. J., Mitterer, H., and McQueen, J. M. (**2011**). "Constraints on the processes responsible for the extrinsic normalization of vowels," Attn., Percept., Psychophys. **73**, 1195–1215.

Sjerps, M. J., and Reinisch, E. (**2015**). "Divide and conquer: How perceptual contrast sensitivity and perceptual learning cooperate in reducing input variation in speech perception," J. Exp. Psychol. **41**, 710–722.

Stilp, C. E., Alexander, J. M., Kiefte, M., and Kluender, K. R. (**2010**). "Auditory color constancy: Calibration to reliable spectral properties across nonspeech context and targets," Attn., Percept., Psychophys. **72**, 470–480.

Stoet, G. (**2010**). "PsyToolkit: A software package for programming psychological experiments using Linux," Behav. Res. Methods **42**(4), 1096–1104.