An Exploration of Bistable Auditory Experiences

Alyssa Rabinowitz

Ramapo College of New Jersey

Abstract

Perception of bistable images, in which a single stimulus can be experienced two ways, has been studied extensively in the visual domain, but less so in the auditory domain. This study presents the Laurel/Yanny recording, a viral video clip of a dictionary pronunciation with two interpretations, as an auditory equivalent of a visual bistable image. It first tested the possibility that people could be made to switch the word they initially perceived in a short-term timeframe by adding contextual sounds as priming stimuli, using both high- and low-order processing. Results showed a substantial percentage of participants switched what they heard, but it was necessary to repeat the study to control for ecological issues. A final experiment replicated a recent study of the Laurel/Yanny phenomenon by using acoustically altered precursors, and found significant differences in the use of a low versus high-frequency precursor, but opposite to the direction of the effect in the first experiment. The main finding was that specific conditions could be created to facilitate certain perceptual experiences. The illusion could be investigated further by focusing on the non-auditory factors that interact with auditory experiences.

An Exploration of Bistable Auditory Experiences

A recent viral internet video emerged in which a recorded word could be perceived in two distinct ways, known to many as the Laurel/Yanny debate. Those defending each word often strongly expressed that their perception was the correct one, and a phenomenon was born. This is similar in many ways to “the dress” phenomenon, which was a viral visual illusion in which a dress could be viewed as blue and black or white and gold, which differed from many other illusions because the individual variation was so strong. This internet trend may have interesting implications in the realm of perceptual psychology, and the following research provides the basis of this study of the Laurel/Yanny auditory illusion.

Wallisch (2017) surveyed 13,417 participants to learn about the perceptual basis of the dress illusion. In the online survey, participants were asked about information regarding their perceptual experience of the dress including the colors they saw, if they switched at any point, and if they perceived the dress to be in shadow. The responses showed that the most influential aspect that determined perceived color was the assumptions of the illumination of the dress in the picture. Also the majority of participants who switched their percept went from white and gold to blue and black (the true colors).

Research by Ward and Scholl (2015) on perceptual switching of visual bistable images used the spinning dancer illusion, in which the dancer can be perceived as spinning clockwise or counter-clockwise. To study the causes of switches in perception, they added line contours at certain times and places to change the orientation of depth and rotation direction of the figure. Participants were exposed to these cues either overtly, or unconsciously. They found that both overt and short unconscious cue could induce switching of the perception of the dancer, and participants were more likely to switch when the cue opposed the percept they were currently experiencing.

Other research on visual bistability focused on the importance of attention on our perception (Dieter, Brascamp, Tadin, & Blake, 2016). They studied binocular rivalry to try to determine the role of attention in which image becomes dominant, and compared the role of attention in binocular rivalry to other forms of visual bistable stimuli. They found that switching involved in binocular rivalry depended on visual attention, but other bistable stimuli like the Necker cube do not require such attention.

A study by Pressnitzer and Hupe (2005) looked at bistability in auditory and visual perception to compare the two. In it, 23 participants were exposed to a visual bistable stimulus of two rectangular shapes moving over a background, and an auditory bistable stimulus of alternating high and low frequency tones. Overall, they found that similar perceptual organization mechanisms are responsible for bistability in both visual and auditory perception. The main difference was the factor of volitional control, or trying to see/hear one percept over the other. This volitional control had a stronger effect in the auditory condition than the visual.

A different auditory illusion has been studied extensively by Deutsch (1987), called the tritone paradox. This is the phenomenon that two tones played one after the other can be perceived as either ascending or descending. In testing four participants by exposing them to a series of tones in this illusion in different keys, she found that the perception of ascending or descending pitches is influenced by where the notes fall on a pitch-class circle, or what key the notes are played in. The internal components of the notes played had an important role in perception. She also found individual differences in the perceptions of this illusion that seemed to be explained by some participants being influenced by the perceived height of the tones, while others were more influenced by the amplitude of the tones' sine waves.

Holt (2006) studied the importance of contextual information in our auditory perception at the neuronal level. In a series of experiments, participants listened to speech and non-speech tone stimuli and reported perceptions of the syllables "ga" and "da" relative to the stimuli. She found evidence that the neighboring sounds had a significant effect on the perception of the syllables. Both the top-down influence of the speech sounds and the bottom-up influence of the tones affected the perception of the targeted speech sounds. This proves that contextual sounds are critical to interpretation of auditory information.

The process of perceptual switching in the auditory modality has been studied in terms of integration and segregation (Denham et al., 2014). Integration refers to grouping sounds into a single stream, whereas segregation refers to separating streams and patterns when perceiving a sound. In one experiment, six participants listened to patterns of 2 tones in different combinations to look at how the tones were either integrated or segregated. They first took three training sessions to be familiarized with the different patterns, and then participated in seven listening sessions all over the course of a month. Participants identified six different patterns of the sequence, and it was found that the ease of identifying a pattern predicted which pattern they heard. Because individuals' responses were consistent when tested again a year later, each participant's perception was a stable characteristic for them. These individual differences in perception can thus be categorized by experiences with multistable auditory stimuli.

Rankin, Sussman, and Rinzel (2015) used the same stimuli of tone pattern differentiation in order to create a model of the neural processes underlying integration and segregation. They tested if, using the model, a change in the difference between the tones' frequencies would affect how long the dominant percept persisted. They found evidence for the model, and the fact that neural dynamics identify a person's perceptual experience of bistability, rather than pre-assigned consciously.

Other research on auditory perception focused on the identification and grouping of pitches, or a form of Gestalt grouping (Mongoven & Carbon, 2017). They tested participants by exposing them to a musical composition and having them identify the symmetry of the notes. The 28 participants were shown different sonifications with note symmetry, partial symmetry, and asymmetry. They found that overall, sonifications with a greater number of notes were more difficult to group, and notes with shorter duration and faster tempo were harder to categorize as well. Accuracy was higher than chance but not perfect, and the participant who scored highest was highly musically trained and could pick apart the notes more efficiently.

The Laurel/Yanny phenomenon was addressed directly for the first time by Bosker (2018), and was published during the course of this study. 532 participants were tested and listened to a 7-point continuum of the Laurel/Yanny stimulus, modulating the frequencies, with 1 sounding the most like Laurel and 7 sounding the most like Yanny. Before each clip was a lead-in stimulus of a telephone number played either with a high-pass filter (as if heard over the phone) consisting of higher frequencies, or a low-pass filter (as if heard through a wall) consisting of lower frequencies. participants then reported if they heard Laurel or Yanny after each step of the continuum was played. He found that the higher frequencies gave more phonetic cues to bias perception toward Yanny, and lower frequencies toward Laurel. In addition, the low-pass filtered precursor phone number led to a higher perception of Yanny responses than the high-pass precursor.

Based on the previous research on visual perception regarding bistability, this study proposed the hypothesis that the Laurel/Yanny illusion is a bistable auditory image that people can be influenced to switch between. We devised two possible methods of manipulation of the context in order to cause people to switch from Laurel to Yanny or vice versa. First, it was proposed that a cognitive priming stimulus, hearing one word repeatedly, would cause participants to hear that word. Secondly, a sensory stimulus of a pitch as priming was predicted to cause participants to switch to hear the word associated with the frequencies of the given pitch.

Experiment 1

Method

Participants. The participants in this study consisted of 46 Ramapo College students, the majority of which were chosen from the subject pool in introduction to psychology classes. A small minority were voluntary participants. Each participant from the subject pool received compensation for their participation in the form of course credit. They were required to sign an informed consent form prior to participating, and data sheets were stored separately from consent forms to maintain anonymity.

Research Design. The experiment was conducted in 2 phases, with each phase testing a separate hypothesis for all participants. The independent variables were the stimuli used as contextual priming manipulations. In the repeated word first phase, the independent variable was either the pronunciation of the word laurel from the Merriam Webster online dictionary, or of the name Yanni on YouTube, depending on the participant’s response. In the second phase, the independent variable was either the high or low frequency pitch. The use of pitches for the second phase of manipulation was based on a spectrogram of the Laurel/Yanny illusion published in an article by the New York Times. This is a visual breakdown mapping the frequencies involved in the bistable stimulus (CITE). That, along with Bosker's recent study (2018), led us to play a high pitch of 40000 Hz to induce switching to Yanny, and a low pitch of 500 Hz to induce switching to Laurel. The dependent variable was what the participant reported hearing in response to the bistable Laurel/Yanny stimulus.

Measures. The dependent measure was the reported perception of the Laurel/Yanny illusion recording. This consisted of a range of answers but was streamlined into the categories of "Laurel," "Yanny," or "other" if it did not resemble either word closely enough. For example, if a participant heard "yammy," that was recorded as "Yanny," and if two words were heard simultaneously but one dominated, the dominant word was recorded.

Procedure. Each participant was tested individually in a research room on the Ramapo College campus. After signing the consent form, the participant sat facing a speaker on the wall, turned away from the experimenter and computer screen from which the stimuli were played. Each recording was played for 30 seconds. The bistable stimulus, the original Laurel/Yanny illusion, was played first, and the participant reported what they heard, and was coded as either Laurel, Yanny, or other. If they heard Laurel, the recording of the repeated word "Yanni" was played, and if they heard Yanny, the repeated word "laurel" was played, both followed immediately by the bistable stimulus again. After this first phase manipulation, the participant was again asked what they heard during the illusion. If they heard Laurel after the first phase, then the 4000 Hz tone was played; if they heard Yanny, the 500 Hz tone was played. Immediately after the tone the bistable stimulus was played again, and the participant reported what they heard after this second phase. Hearing the opposite word as was heard the preceding time would denote the participant as a “switcher” and dictated which repeated word or pitch prime he or she was played (see Figure 2).

Results

The results consisted of proportional analysis of participants who switched what they heard after being exposed to either or both of the two phases of stimuli. Out of the 46 participants, 10 participants switched at some point (21.7%). Of those 10 “switchers,” 5 switched only after the first phase, 1 switched only after the second phase, and 4 switched twice. Of the 10 participants who switched, 80% heard Yanny dominantly first. Put differently, 9 participants switched after the first phase in response to the cognitive prime (repeated word), of which 5 switched back after the second pitch priming phase. Only 1 participant switched only after the pitch prime. Of the 36 participants who did not switch, 12 heard Yanny only, while 24 heard Laurel only.

Discussion

The findings of the current study showed some, but minimal, support for our hypotheses. They were generally consistent with the research on both visual and auditory bistability, as well as the trends for this illusion specifically. It is the first study of our knowledge to create specific conditions under which switching to a different percept was more likely, though the effect was minor. The breakdown of non-switchers, with twice the amount of participants hearing Laurel as Yanny, is consistent with statistics on general perception of the illusion. This experiment proves a previously unknown trait that may be associated more with hearing Yanny (the ability to switch), but would need repeated trials to confirm this.

Our results were consistent with Bosker (2018) in that individuals can be manipulated to hear Laurel where they previously heard Yanny. We also found that the majority of switchers started with Yanny and switched to Laurel. We found a lower percentage of participants who were affected by the contextual stimuli used here, but this could have been due to our smaller sample or a number of factors discussed below.

Our results of the switchers going specifically from Yanny to Laurel (the true word) is similar to the results found in the study of “the dress” by Wallisch (2017). In his study, most switchers went from white and gold to blue and black (the true colors), and in ours most participants went from Yanny to the true word, Laurel. The main difference is that most people perceive the Laurel/Yanny stimulus to be the true word, Laurel, whereas the majority of people perceive the dress as white and gold. Also, the importance of the context seems to be translated fairly well to auditory processing, as seen in the current study.

The spinning dancer illusion studied by Ward and Scholl (2015) provided interesting similarities to the Laurel/Yanny stimulus. Both seem to have infrequent switching, unlike many other bistable images. Still, participants could be made to switch even unconsciously, and this may parallel the switching in Laurel/Yanny. The unconscious cues that resulted in immediate changes in perception may be related to the participants who we found to be able to switch short term in this experiment. Dieter, Brascamp, Tadin, and Blake (2016) helped to understand why bistability may still exist even with a lack of focused attention. Although they did not consider auditory stimuli, this fact can be applied to our study because it is clear that even when people do not focus on hearing one thing over the other, one word still persists without conscious effort, and it doesn’t just become a disorganized jumble of sound.

The main finding of Pressnitzer and Hupé (2015) that bistability exists in the auditory realm similarly to visual perception influenced our attempt to manipulate the Laurel/Yanny illusion as a bistable stimulus to begin with. Though not studied here directly, it does not appear that there is a volitional component of the illusion used, that would be interesting to consider in subsequent experiments. The study of another auditory illusion (Deutsch, 1987) helped to provide a basis for our interpretation of the Laurel/Yanny illusion as having individual differences that could be categorized. We wished to add the additional component of a manipulation that could help explain the different influences for each word, similar to her different components of a pitch that influenced perception of the tritone paradox.

The study by Holt (2006) influenced the two stimuli used in this study: the top-down influence of the repeated pure word, and the bottom-up influence of the tones. We similarly found that the linguistic speech sounds had a contextual influence on the perception of Laurel/Yanny, but the bottom-up pitch influence was not as successful. The possible explanations for this follow.

Limitations. There are several factors relating to controls and test conditions that may have prevented a larger effect from being observed in the results. A major downfall of this study was not accounting for the spectrum of responses that did not fall definitively into “Laurel” or “Yanny.” The one participant who only switched as a result of the pitch prime did not experience a very clear switch from one word to the other, for example. In the future, the range of responses could be categorized more precisely, or participants should be prompted with the two possible responses in advance to eliminate this problem altogether.

The general lesser success of the second phase stimulus, the pitch prime, can possibly be explained by the process of neuronal adaptation. Because at a low level of processing the brain responds best to novel auditory stimuli (Holt, 2006), it may have been better to switch the way the pitches were used. This would mean playing the low frequency associated with Laurel if the participant had heard Laurel to have them switch to Yanny, rather than play the high frequency associated with the target word Yanny.

Another limitation was the inconsistencies and distractions within the test environment. There might have been distracting sounds, and the room was not 100% soundproof. There were slight differences in the volume levels and instructions by experimenters that may have led to less consistent results. In the future, it would be better to have participants use headphones and have more controlled test conditions.

In a future experiment, it would be useful to differentiate between participants who were naive to the Laurel/Yanny illusion or not. This is because there might be another individual difference between people who have heard it before and have a preconceived notion of what their perception will be, and those who do not know what the options even are. It is possible some participants were strong in their belief that they expected to hear something, and this cognitive process could play a role, especially taking into account research on the influence of volitional control in perception.

One thing we cannot be sure of from this study is if the participants who switched their perceptions following our manipulations would have switched regardless of the stimuli presented. Since there was no control group that listened to the stimulus three times without the two priming precursors, we can only speculate that our variables played a role. Still, the fact that a decent proportion of listeners were were able to switch in such a short time frame is a significant finding in itself.

Experiment 2

The purpose of the second experiment was to replicate the first, but address some of the limitations of the previous study. It was revised to control environmental conditions like speaker volume and inter-rater reliability, and take into account the prior experience of the participants. This could potentially help determine the role of expectations in perceiving Laurel or Yanny or at least show a pattern of individual characteristics that promotes or prevents perceptual switching.

Method

Participants. The participants consisted of 63 Introduction to Psychology students at Ramapo College, and were sampled by convenience. One class of 33 students constituted the first experimental condition, while another class of 30 made up the second experimental condition. The experiments were run at the beginning of each class.

Research Design. The participants were tested in a group setting, and the primes played were not dependent on individual responses, but rather were pre-selected by the experimenter. For the first group, the primes were the repeated word “Yanni” and the 4000 Hz tone. Participants in the second condition heard the repeated word “laurel” and the 500 Hz tone.

Measures. The participants again used self-report measures to record the word they heard. This self-report asked the additional question of if they had previous experience hearing the clip.

Procedure. The participants of each group were tested as a whole, rather than individually. The procedure was the same as Experiment 1, but without asking for a verbal response after each time the bistable clip was played. Instead participants were told to record what that had heard after each of the three times hearing the original Laurel/Yanny clip. The first group was played the bistable stimulus, recorded a response, then was played the repeated word “Yanni” followed by the bistable, recorded a response, then was played the high pitch followed by the bistable, and recorded a response. In the second class, the repeated word “laurel” and the low pitch replaced the repeated “Yanni” and the high pitch (see Figure 3). After the experiment ended, participants were prompted to write down if they had heard the bistable stimulus recording prior to participating in the study.

Results

The results showed 15.87% of all participants, five in each class, switched what they heard at least once. In the first group, two participants switched after the repeated word prime from Yanny to Laurel, one switched after the high pitch prime only from hearing both to hearing Yanny, and two switched twice, both from Yanny, then to both, then to Laurel. In the second group, after hearing the repeated “laurel” prime, four participants switched from Yanny to Laurel, and one switched from Yanny to a different word, “yeilly.” None switched after the low pitch prime in that group.

Upon the initial exposure to the bistable stimulus in group 1, 10 participants heard Laurel, 18 heard Yanny, and 5 were other. In group 2, 11 participants heard Laurel, 15 heard Yanny, and 4 were other. Overall, one third of participants initially heard Laurel, and just over half heard Yanny initially.

Of all the non-switchers, 21 participants heard Laurel only, 25 participants heard Yanny only, and six participants heard something other than Laurel or Yanny. Thirteen participants (20.63%) total had no prior exposure to the bistable stimulus, and 38 (60.32%) did. Two did not respond. Of the 13 who had no prior exposure, 1 switched; of the 38 who had prior exposure, 7 switched. Of all those who switched, two participants had no previous exposure, six participants did, and two failed to respond. All 10 “switchers” started out hearing either Yanny or both words simultaneously.

Discussion

This experiment made it possible to see what the priming stimuli used in Experiment 1 had on those who heard Laurel and those who heard Yanny. It also gave the chance to see if a low pitch prime would actually lead to more Yanny responses, opposite of what was used in the first experiment. Because in Condition 1 (“Yanni” repeated as the first prime) all the switchers were those who initially heard Yanny, and no one who heard Laurel then switched to Yanny, the efficacy of that prime is questionable. The directions of the switches were not consistent, especially since none of the participants in the first class who initially heard Laurel switched to Yanny. This may point to ineffectiveness of the primes to cause switching in a specific direction, contrary to the effect assumed from the Experiment 1 results.

The prior exposure information presented a tentative pattern: 18.42% of those who reportedly had experience with the stimulus switched, and 7.69% switched of those who did not. The possibility of past experience acting as a reliable predictor of switching would have to be assessed in further studies.

Limitations. Several factors could have contributed to the results being inconsistent with the first experiment. Ecological factors relating to the group setting rather than individual tests could have played a role, including distractions, communication between participants, or general inattentiveness. In addition, the participants did not sign up to take part but rather were surprised with the experiment, and although they had the option to opt out after reading the informed consent, they may have felt group pressure to participate. Also, because of the nature of the self-report measure, responses were often vague and possibly could have been misinterpreted. The coding of responses was similarly limited, and did not allow for the response of switching that occurred within the 30 seconds of exposure to the bistable stimulus. We also did not differentiate between the range of “other” responses. For example, “Yanny then Laurel” or “mostly Yanny” was coded as Yanny, but “Laurel and Yanny,” “both,” or “daily” was coded as other.

This experiment was an important addition in pointing out the inconsistencies in the primes devised here to cause switching in a certain direction, emphasizing the need to further validate them. It did help point to a general pattern of who is more likely to switch, regardless of contextual sounds – those who initially hear Yanny. It would be useful to test if the order of primes played makes a difference in switching responses as well.

The high/low frequency pitch primes seemed to be the most unanswered question in terms of the effect produced, if any. The absence of a control group here was mainly due to the shift in experimental focus to better answer this question. For this reason, it seemed a good option to reproduce Bosker’s (2018) study, which obtained significant results by using a different pitch-manipulated precursor to the bistable stimulus: a telephone number.

Experiment 3

This experiment was intended to replicate the results found by Bosker (2018), that emphasizing higher pitches within the bistable clip leads to the perception of Yanny, and that people can be made to hear Yanny after first previously hearing Laurel. This was accomplished by using a higher-pitch and lower-pitch recording of a phone number immediately before hearing the bistable stimulus, which was broken into a continuum of 7 steps (step 1 sounding the most Laurel-like and step 7 sounding the most Yanny-like. It was then hypothesized that by using repeated exposure to the precursors each time (three times as opposed to one), this effect of the precursor recordings could be magnified.

Method

Participants. The participants of this study were 54 students in perceptual psychology classes at Ramapo College, in a convenience sample. One class of 30 students constituted the first experimental condition, and one participant was excluded due to irrelevant responses, leaving 29 total participants in group 1. Another class of 25 students made up the second experimental condition, group 2. The experiments were again run at the beginning of each class.

Research Design. Both experimental groups had the same design, with the only difference being the repeated precursor in the second group. The 7 steps of the continuum were paired with both the high-pitch precursor and the low-pitch precursor, leading to a total of 14 experimental clips. This was repeated three times in random order, leading to 42 distinct responses.

Measures. The dependent measure was the self-reported word heard for all three trials of the 14 clips.

Procedure. The participants were told as a class they would be participating in a senior psychology student’s research study, and were then given informed consent forms to fill out. The reporting sheets were distributed, and then brief instructions of the experiment were given. They were told they would hear a series of recordings, in which a telephone number would be paired with a word, and to write down the final word they heard each time. In the first class, the low-pitch phone number was played once preceding each of the 7 steps on the Laurel/Yanny continuum, and the high-pitch phone number was played once preceding each step. These 14 responses constituted one trial, and three trials were given (the order of the steps was manually randomized before the experiment). In the second condition, the same procedure took place with the only difference being that the precursor phone number was played three times in a row preceding each Laurel/Yanny recording. After the experiment concluded, the class was asked if anyone had not heard the illusion before, and were debriefed on the purpose of the study.

Results

First, he results of each condition are displayed as a separate line graph (Figs. 4 and 5). The slope of the line represents the effect of the continuum steps, going from the the lower pitches of the bistable clip emphasizes to the higher pitches emphasized. For both conditions, the total percentage of Yanny results for each step, both for the high and low precursors, was plotted as a point. The lower-pitch precursor produced more Yanny results in a majority of the steps in the first condition, and consistently in the second condition. For the low-pass graph, step 6 had a significant outlier from one trial disrupting the otherwise consistent positive trend of the results. For statistical analysis, a paired-sample t-test was conducted to compare the average results of each class between the low- and high-pitch precursor for each of the 7 steps, using a .05 alpha level. For the 1x condition, no significant effect on Yanny responses was found between the precursors ($t(6) = .72$, $p = .25$). For the 3x condition, there was a significant difference between the number of Yanny responses from the low and high precursors ($t(6) = 2.63$, $p = .01$).

Then, a second set of line graphs represent the results of each precursor, with both conditions represented on a single graph. These compare the high-pitch precursor for both conditions, and the low-pitch precursors for both conditions (Figs. 6 and 7). These results show that the 3-times repeated precursor condition produced significantly fewer Yanny results than the single precursor condition, contrary to the hypothesis. A paired-sample t-test indicated significant differences in Yanny responses for the low precursor between the 1x and 3x repetition, ($t(6) = 2.07$, $p = .04$), and significance for the high precursor between the two groups ($t(6) = 1.96$, $p = .04$).

Overall, 40.7% of all participants were perceptually stable, reporting the same word over 90% of the time. In group 1, 44% of participants were stable in hearing Laurel, and 4% were stable in hearing Yanny. In group 2, 34.5% of participants heard Laurel consistently.

Discussion

One difference between these results and Bosker's (2018) is that the percentage of Yanny responses does not increase with each step, except for the low-pass precursor in Condition 2. However, the more pertinent variables to this study was the effect of the precursor, and if repeated exposure to the precursors would amplify the results. The effect of the low and high-pitch precursors was not replicated in the 1x condition, though playing the precursor once preceding each step was effective in the original results (Bosker, 2018). By repeating the precursors three times before each continuum step, we were able to achieve the significant results of the two precursor's differential effects, with the low-pitch precursor producing more Yanny responses.

It seems likely that there was experimenter error while administering the experiment in the first group. The clips may have been played in a different order than was recorded as played, and this would explain the lack of steady increase of Yanny as the steps get higher. In group 1, the data had two most noticeable outliers of Yanny responses in the last trial for step 1with the low-pitch precursor and step 6 with the high-pitch precursor. Thus, it seems plausible that these two may have been switched in the data recording, because the graph would otherwise follow a trend consistent with group 2. This change would make the difference between precursors in group 1 significant, but would make the difference between the two groups for the low precursor no longer significant. This cannot be confirmed, but it is a compelling explanation. I would repeat the study if the research could be extended to eliminate this concern, and to test the reliability of the materials.

Repeated exposure to the precursor did produce a significant effect compared to playing it once, but in the opposite direction as predicted. Instead of repeated precursor exposure leading to greater Yanny responses, it led to fewer. A potential explanation is the process of habituation, in which a perceptual response diminished after repeated exposure to a stimulus. It is an explanation similar to why playing higher frequencies associated with Yanny is not as effective to induce hearing Yanny than using the frequencies associated with the opposite word.

Another major difference between this experiment and the one it was modeled after was that the percentage of Yanny results by step 7 only got to around 50% for each condition, compared to the Bosker (2018) study which found about 90% Yanny responses for step 7. This could be due to chance and that the smaller sample size used here had so many consistent Laurel-hearers, or other environmental factors. These overall different effects of the continuum steps from Bosker (2018) does not exclude the importance of the effect of the precursor. In addition, the Bosker study utilized five trials compared to the three used here, and there were almost 10 times more participants in that study. This means that the scores Bosker found would have averaged out the total percentages, while the results here are subject to more fluctuation and influence of individual responses.

Another key difference to help account for the differences between the current experiment and Bosker (2018) is the portion of participants who were perceptually stable. These are the people who heard the same word more than 90% of the time and were not affected by any of the experiment’s manipulations. Almost double the proportion of participants were perceptually consistent in this study as in the original Bosker study, which found about 20%

stability. In addition, these people almost all were stable in hearing Laurel, rather than Yanny. This is consistent with our previous findings that those who hear Laurel are less likely to have freely bistable perception.

Bosker (2018) used the theory of neuronal adaptation to explain why the low-pitch precursor led to increased reports of Yanny, which is associated with the higher frequencies of the original bistable recording. It is still just a theory, though, so it is not impossible that this concept was simply applied to his results after the fact, as an explanation of hearing the word opposite the pitch priming. In the same way, this theory could tentatively explain the significance change caused by the 3x repeated precursor priming in this study.

It is an interesting question why the use of targeted frequencies as priming stimuli was successful here, but less so in Experiment 1. The use of the phone number as the precursor was perhaps a better low-level prime than the straight tone that was used in Experiment 1. The linguistic, human voice aspect could have made it more effective in altering perception. This would help explain the relative lack of success seen with the frequency prime used in our first study, opposed to the precursors here that are filtered to emphasize lower or higher frequencies.

Limitations. One limitation of the current study is the potential participant differences in choosing perceptual psychology students as opposed to introductory students. It is possible that having more experience in the subject could influence results, though it is likely not major. All of the students in this experiment responded that they had heard the bistable stimulus before, and this is a significant change from the Intro students who were more split with their prior experience. This would be useful to compare in a future study. The same group ecological effects as Experiment 2, such as peer influence and other distractions, were a possible limitation as well.

It should also be noted that the two groups were tested in different classrooms, though they both had the same speaker equipment (two wall-mounted JBL Control 25 speakers). There is a chance that there was a difference in volume or acoustics of the different rooms. This should be controlled in a future replication.

Future research would address the specific procedural differences between this study and Bosker's (2018). A larger quantity and variety of non-psychology students should be sampled and assigned randomly to address any participant differences. It would be interesting to see if the results would mirror the original ones more if the test was administered online, without control of the speakers and headphone use of participants as it was done in the Bosker study. The trials were also not fully randomized, just the order of the continuum steps were. The same low or high pitch precursor was played seven times in a row, and whether the next 7 would be high or low was randomly picked until three trials were completed. More trials could also make a difference in averaging out responses.

Conclusions

The Laurel/Yanny illusion has been shown to be a phenomenon comparable to a number of visual bistable images like the spinning dancer and old/young woman illusions, because it has been shown that switching does occur in the short term – though it is not the majority of listeners. The results of these experiments point to an ability to change what we perceive as a result of a simple short-term contextual change. It also seems likely that by repeating these studies with more participants and addressing experimenter, selection, and ecological limitations, a better trend could be created revealing differences in experience.

One area that this study did not go into depth in was the role of attention on auditory perception. In the visual domain, this has been a frequently observed phenomenon in regard to optical illusions (Ward & Scholl, 2015; Dieter, Brascamp, Tadin, & Blake, 2016). What we attend to is a major part of these bistable experiences. It is clear that internally, by masking certain frequencies so that we attend to others, we are more likely to hear one word over the other. Externally, there are a wider variety of factors that contribute to our perceptions, many of which are not easily controlled in experimental conditions.

The different findings of the present study are consistent with findings on the importance of individual differences in multistable perception (Denham, et al., 2014). For most of our participants, there may have been a stable characteristic of their perception that prevented the switching based on what was initially heard. Mongoven and Carbon (2017) noted the influence of the experience and education of individuals that may have led to some differences between perceptions of specific components of music. Since having musical training played a role in their study, this could help explain the variation between individual participants in tests of bistable stimuli, including the auditory Laurel/Yanny stimulus used here.

There could also be a range of other priming stimuli that could bias perception, not just sounds. The McGurk effect shows that competing auditory and visual information interact and can produce entirely new perceptions. The visual information of lips mouthing “ga” and audio information of the sound “ba” leads people to perceive the syllable “da” (McGurk & MacDonald, 1976). It has also been shown that cross-modal perception can greatly influence the perception in a single domain (Shinohara, Yamauchi, Kawahara, & Tanaka, 2016). They found that sound can have symbolic meaning, and can be expressed as shapes and dynamic movement as well. More round, flowing movement of lines is associated with softer sounds and words opening with vowels, while angular, sharp moving lines are associated with harsher consonant sounds.

This research began with a hunch that there was more to the Laurel/Yanny illusion than just individual differences in hearing. Much of the debate surrounding the viral clip was, like with “the dress,” how unwavering people’s opinions were about what they perceived. This was largely an attempt to dismantle that assumption. We sought to explore how our hearing can be manipulated by outside sounds through high and low-level processing, and this confirmed the initial hunch as correct, though refinement is certainly needed.

There are things that can be done in the future to show greater success in the manipulations used, but there are a variety of other manipulations that could be explored. It is possible that higher, cognitive processes could play a more significant role, so a study just testing different ways to make someone think of the word laurel or Yanny might be enlightening. As for lower level auditory processing, this research has given a new lens through which to examine neuronal adaptation and habituation. There is no reason that a binary should exist such that neural processing and higher-order cognition cannot play interacting roles in multistable auditory perception, and that interplay should be the focus moving forward.

References

Bosker, H. R. (2018). Putting Laurel and Yanny in context. *The Journal of the Acoustical Society of America*, *144*(6). doi:10.1121/1.5070144

Denham, S., Bõhm, T. M., Bendixen, A., Szalárdy, O., Kocsis, Z., Mill, R., & Winkler, I. (2014). Stable individual characteristics in the perception of multiple embedded patterns in multistable auditory stimuli. *Frontiers in Neuroscience, 8*. doi:10.3389/fnins.2014.00025

Deutsch, D. (1987). The tritone paradox: Effects of spectral variables. *Perception & Psychophysics, 41*(6), 563–575. https://doi.org/10.3758/BF03210490

Dieter, K. C., Brascamp, J., Tadin, D., & Blake, R. (2016). Does visual attention drive the dynamics of bistable perception? *Attention, Perception, & Psychophysics,78*(7), 1861-1873. doi:10.3758/s13414-016-1143-2

Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *The Journal of the Acoustical Society of America, 120*(5), 2801-2817. doi:10.1121/1.2354071

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748. doi:10.1038/264746a0

Mongoven, C., & Carbon, C.-C. (2017). Acoustic Gestalt: On the perceptibility of melodic symmetry. *Musicae Scientiae*, *21*(1), 41–59. https://doi.org/10.1177/1029864916637116

Pressnitzer, D., & Hupé, J. M. (2005). Is auditory streaming a bistable percept? *Acta Acustica united with Acustica, 91*(S1), S102.

Rankin, J., Sussman, E., & Rinzel, J. (2015). Neuromechanistic Model of Auditory Bistability. *PLOS Computational Biology,11*(11). doi:10.1371/journal.pcbi.1004555

Wallisch, P. (2017). Illumination assumptions account for individual differences in the perceptual interpretation of a profoundly ambiguous stimulus in the color domain: “The dress.” *Journal of Vision*, *17*(4). https://doi.org/10.1167/17.4.5

Ward, E. J., & Scholl, B. J. (2015). Stochastic or systematic? Seemingly random perceptual switching in bistable events triggered by transient unconscious cues. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 929–939. https://doi.org/10.1037/a0038709

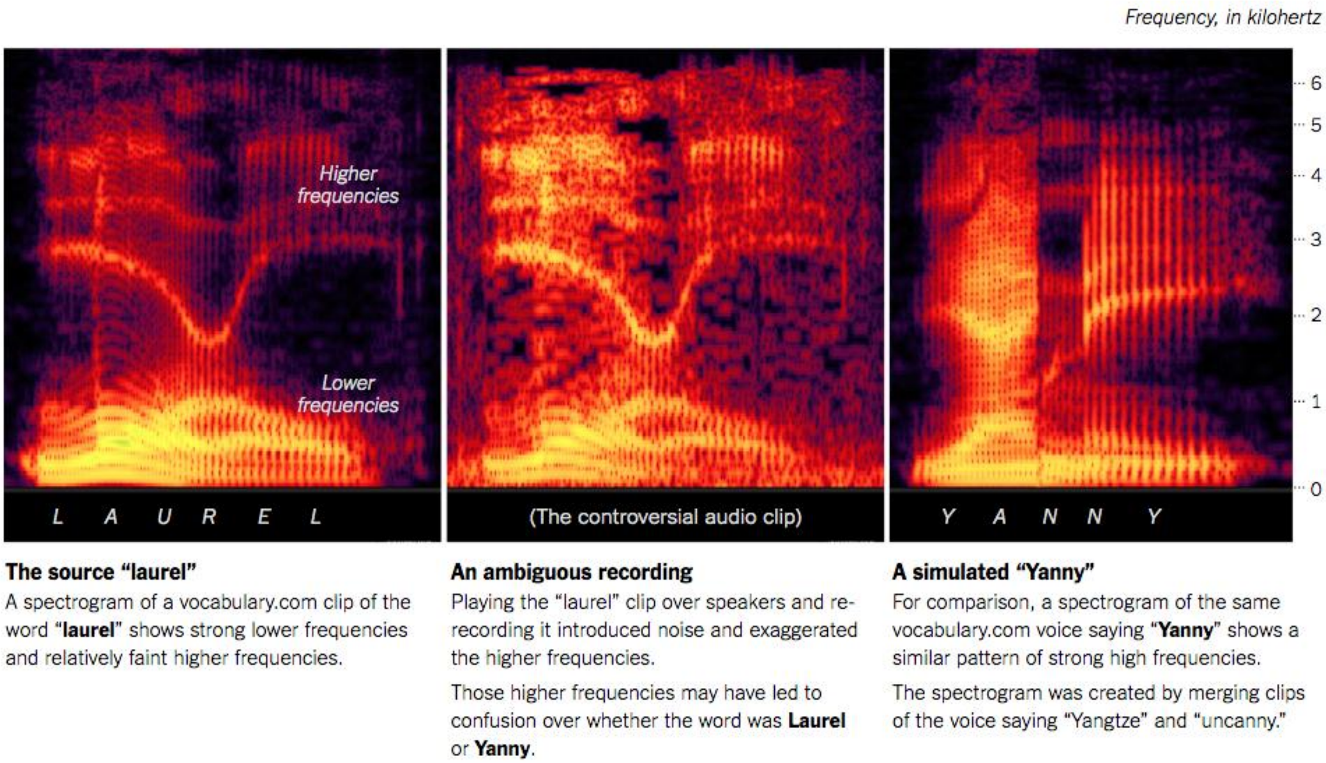

Figure 1. Spectrogram mapping the frequencies associated with the illusion, published by the New York Times (2018).

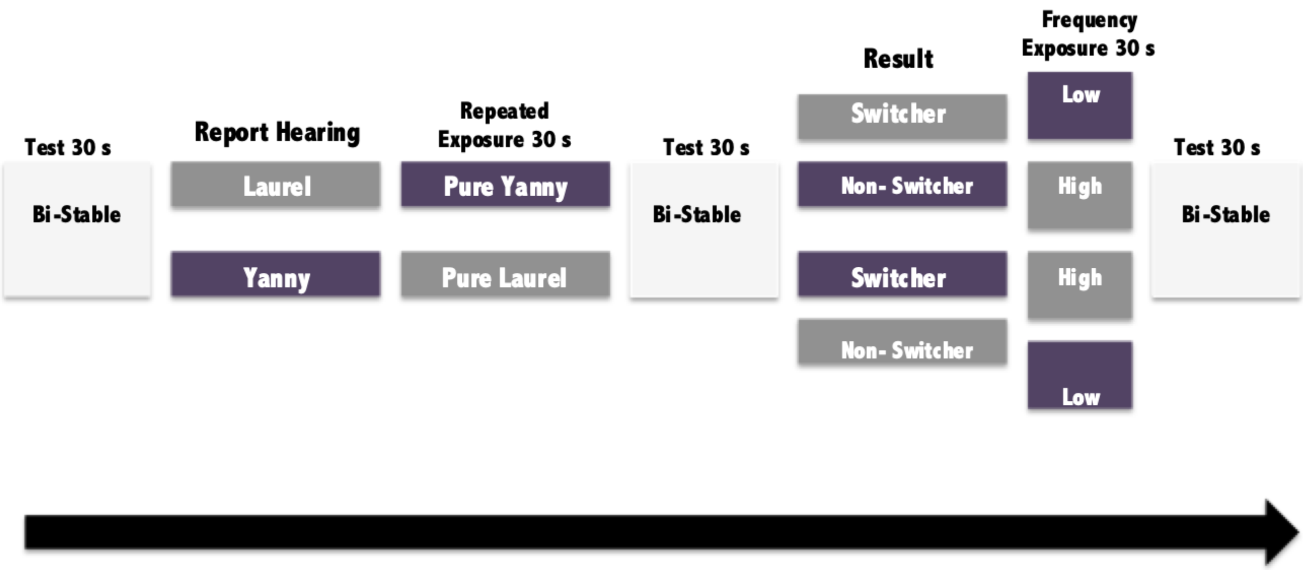

Figure 2. Sequence of Experiment 1 for each participant.

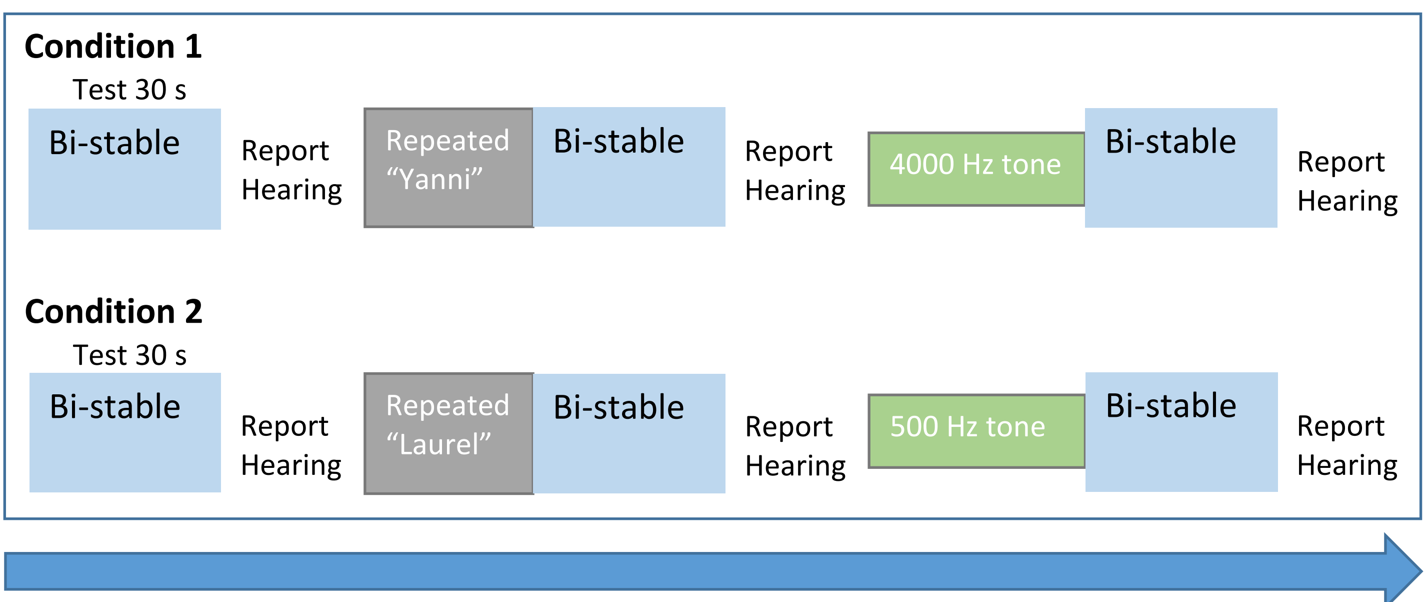

Figure 3. Sequence of Experiment 2.

Condition 1 sequence was heard by the whole first group, and Condition 2 was heard by the whole second group.

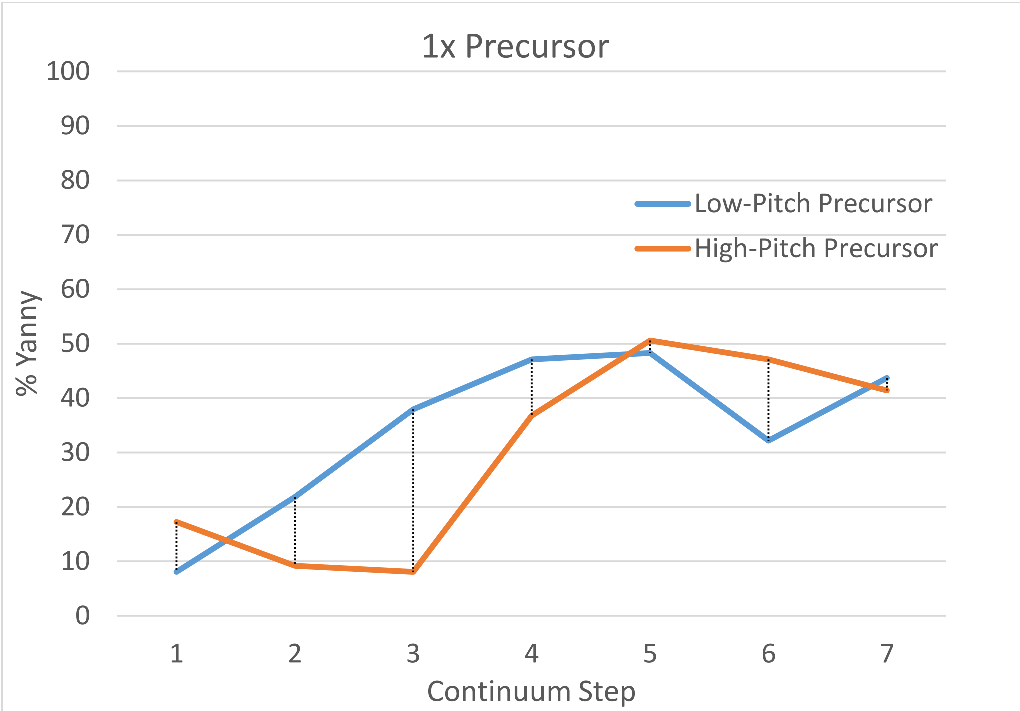

Figure 4. Experiment 3, condition 1, showing results of both pitched precursors within the first experimental group.

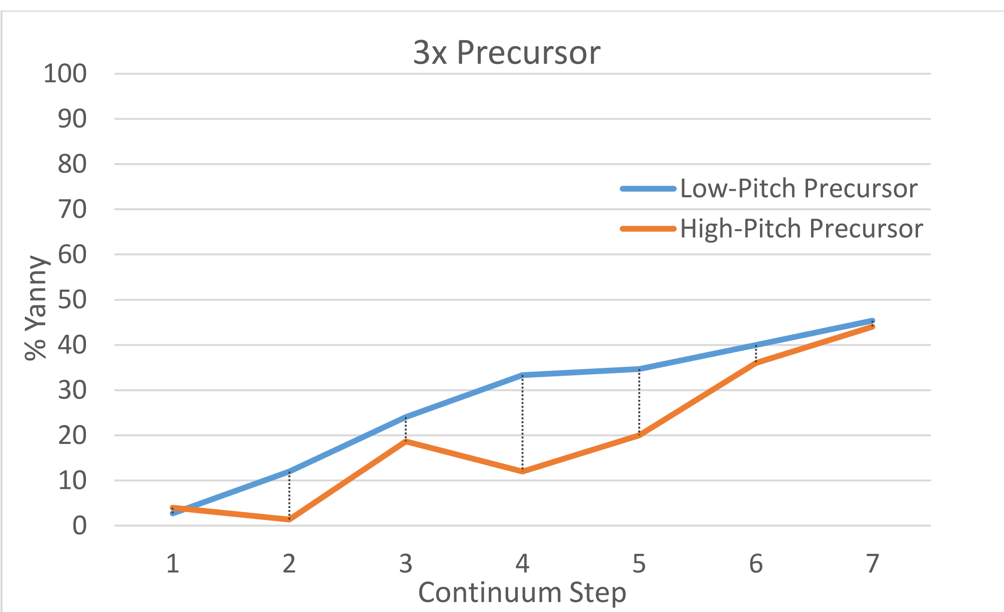

Figure 5. Experiment 3, condition 2, showing results of both pitched precursors within the second experimental group.

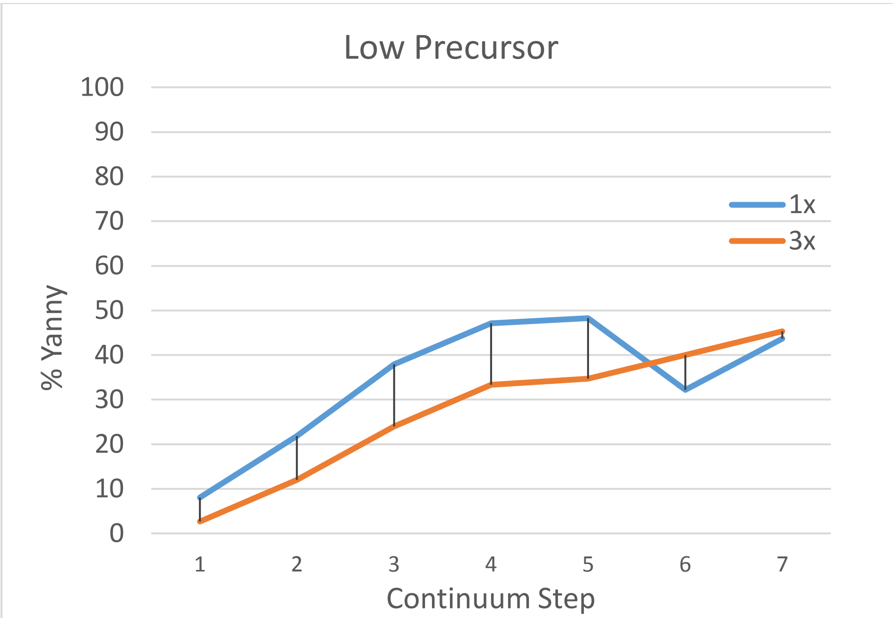

Figure 6. Experiment 3, comparing the effects of low-pass precursor between experimental groups.

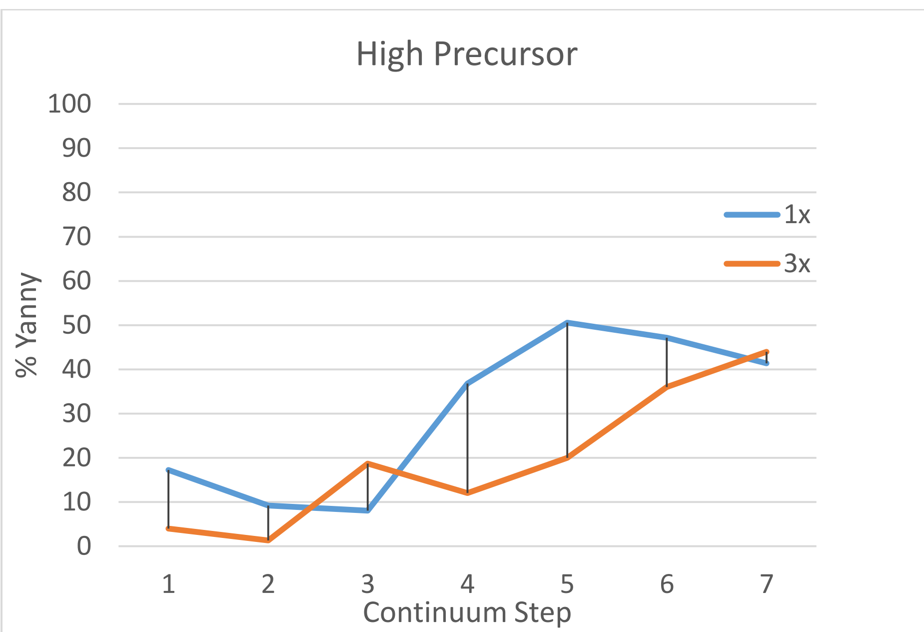

Figure 7. Experiment 3, comparing the effects of high-pass precursor between experimental groups.