

Perceptual hysteresis in the judgment of auditory pitch shift

Claire Chambers · Daniel Pressnitzer

Published online: 30 May 2014
© Psychonomic Society, Inc. 2014

Abstract *Perceptual hysteresis* can be defined as the enduring influence of the recent past on current perception. Here, hysteresis was investigated in a basic auditory task: pitch comparisons between successive tones. On each trial, listeners were presented with pairs of tones and asked to report the direction of subjective pitch shift, as either “up” or “down.” All tones were complexes known as *Shepard tones* (Shepard, 1964), which comprise several frequency components at octave multiples of a base frequency. The results showed that perceptual judgments were determined both by stimulus-related factors (the interval ratio between the base frequencies within a pair) and by recent context (the intervals in the two previous trials). When tones were presented in ordered sequences, for which the frequency interval between tones was varied in a progressive manner, strong hysteresis was found. In particular, ambiguous stimuli that led to equal probabilities of “up” and “down” responses within a randomized context were almost fully determined within an ordered context. Moreover, hysteresis did not act on the direction of the reported pitch shift, but rather on the perceptual representation of each tone. Thus, hysteresis could be observed within sequences in which listeners varied between “up” and “down” responses, enabling us to largely rule out confounds related to response bias. The strength of the perceptual hysteresis observed suggests that the ongoing context may have a substantial

influence on fundamental aspects of auditory perception, such as how we perceive the changes in pitch between successive sounds.

Keywords Adaptation · Aftereffects · Hearing · Psychoacoustics

Sometimes, presenting the same physical stimulus twice results in different perceptual outcomes. This is obviously true around threshold, but this can also occur above threshold, in the perception of stimuli that have more than one plausible perceptual interpretation. Such ambiguous stimuli have been extensively used in the study of visual multistability, where prolonged exposure leads to spontaneous alternations of perception in the mind of the observer (see Leopold & Logothetis, 1999, for a review). The perception of ambiguous visual stimuli has also been shown to be prone to context effects (Hock, Kelso, & Schöner, 1993; Kanai & Verstraten, 2005; Leopold, Wilke, Maier, & Logothetis, 2002; Maloney, Dal Martello, Sahm, & Spillmann, 2005; Noest, van Ee, Nijs, & van Wezel, 2007). In all of these studies, the perception of an ambiguous stimulus was strongly modulated by the recent history of stimulation and/or perception. This occurs because ambiguity serves to highlight general-purpose contextual processes: If nothing in the stimulus favors one interpretation over others, then the context may become the decisive factor. Ambiguous stimuli may thus be a useful experimental tool to characterize contextual processing, without the competing influence of stimulus-related cues.

To date, auditory science has made less use of experimental paradigms relying on ambiguous stimuli than vision science. That being said, instances of auditory multistability do exist, most notably related to auditory scene analysis (see Schwartz, Grimault, Hupé, Moore, & Pressnitzer, 2012, for a review). Context effects in ambiguous auditory stimuli have also been

C. Chambers · D. Pressnitzer
Laboratoire des Systèmes Perceptifs, CNRS UMR 8248, 29 rue
d’Ulm, 75005 Paris, France

C. Chambers (✉) · D. Pressnitzer
Département d’Etudes Cognitives, Ecole Normale Supérieure, 29 rue
d’Ulm, 75005 Paris, France
e-mail: claire.chambers@ens.fr

reported (Holt, 2005, 2006; Huang & Holt, 2012; Laing, Liu, Lotto, & Holt, 2012; Snyder, Carter, Hannon, & Alain, 2009; Snyder, Carter, Lee, Hannon, & Alain, 2008). The aforementioned studies used tasks related to presumably high-level constructs, such as perceptual organization (Snyder et al., 2009; Snyder et al., 2008) or speech phoneme classification (Holt, 2005, 2006; Huang & Holt, 2012; Laing et al., 2012). Here, we aimed to use ambiguous stimuli to investigate the influence of context on a more basic auditory task—namely, comparing the pitch of two successive tones.

The ambiguous stimuli consisted of Shepard tones (Shepard, 1964). A Shepard tone contains several frequency components, all with an octave relation to a single base frequency, F_b . All adjacent components (F_b , $2F_b$, $4F_b$, etc.) have the same ratio, and are thus evenly spaced on a log-frequency scale (Fig. 1). When two Shepard tones are presented in succession, listeners tend to report a pitch shift that corresponds to the shortest distance between the components, on the log-frequency scale (Shepard, 1964; see Fig. 1A). Importantly for our purposes, an interval of a half-octave (the musical “tritone,” or six semitones; see Fig. 1B) produces equal distances for upward and downward frequency shifts. In accordance with this physical ambiguity, subjective judgments are split between “up” and “down” reports (Shepard, 1964). A number of studies have shown that individual listeners can display stable, and sometimes very large, idiosyncratic biases in their perceptions of such stimuli, depending both on the listener’s linguistic background and on stimulus properties such as the pitch class of the notes forming the interval (Deutsch, 1987, 2013; Deutsch, Moore, & Dolson, 1986; Ragozzine & Deutsch, 1994).

In addition to the study of long-term idiosyncratic biases, researchers have also attempted to probe the effect of immediate context in the perception of Shepard tone pairs. Repp (1997, Exp. 3) observed that presenting a single Shepard tone before an ambiguous Shepard test tone-pair could influence the reported pitch shift for the test pair. Listeners appeared to minimize the pitch range of the reported shifts. For instance, if listeners reported a “down” shift between the context and the first tone of the test pair, the shift within the test pair was more likely to be perceived as an “up” shift. In a subsequent study, Dawe, Platt, and Welsh (1998) attempted to induce adaptation of “up” and “down” shifts by presenting sequences of Shepard tone pairs with a dominant interpretation (e.g., “up”) and measuring the pitch shift for an ambiguous test pair (which should then be biased toward “down,” by contrast). They reported a weak and unreliable effect, where either contrastive or assimilative biases were observed depending on the pitch class of the ambiguous test pair. Repp and Thompson (2010) revisited context effects by using nonambiguous tones as their context stimuli. They found no robust effect of context, and concluded that the perception of Shepard tone pairs was largely invariant to context.

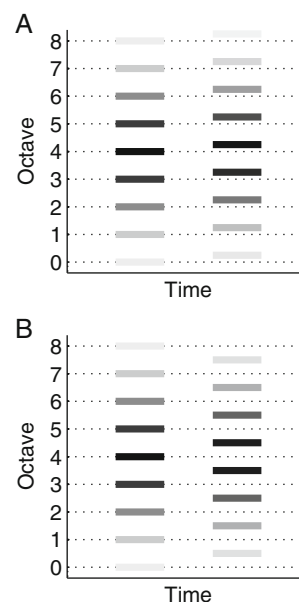


Fig. 1 Schematic spectrograms of two Shepard tone pairs, with frequency on a logarithmic scale over time. Grayscale represents amplitude. (A) Three-semitone interval, for which the dominant percept is “up.” (B) Six-semitone interval, for which the direction of shift is ambiguous

One experiment with a different paradigm led to a sizeable context effect. Giangrande, Tuller, and Kelso (2003) measured hysteresis in the perception of Shepard tones. Sequences of tone pairs were presented, and listeners reported the direction of the subjective pitch shifts. The F_b of the first tone in each pair, the *standard tone*, was fixed across trials, while the F_b of the second tone, the *comparison tone*, was varied. In some experimental conditions, the interval between the standard and comparison was regularly increased between successive trials, in steps of one semitone, starting from a nonambiguous interval. In other conditions, it was regularly decreased. Giangrande et al. observed that the initial responses in the sequences were maintained in the parameter range usually associated with perceptual ambiguity, resulting in sizeable hysteresis in the pattern of responses.

However, some methodological details complicate the interpretation of the findings of Giangrande et al. (2003). As was emphasized by Hock et al. (1993), hysteresis in response patterns may be either perceptual or due to response bias. When conditions give rise to ambiguity, and hence to uncertainty, one possible strategy is for observers to maintain their previous response. In the experiment of Giangrande et al., no measures were taken to dissociate this strategy from hysteresis in perception. Moreover, even if their findings did reflect perceptual hysteresis, the origins of the effect are unclear. It may have been due to a tendency to hear pitch shifts in the same direction on consecutive trials (“up,” “up,” etc.), or alternatively, it could have been a bias in the perceptual representation of each tone (“standard lower,” “standard lower,” etc.).

In the following experiment, we revisited the hysteresis paradigm of Giangrande et al. (2003), introducing a simple methodological modification: We randomized the order of presentation of the standard and comparison tones within trials. As we will explain, this removed the possibility of response hysteresis, while dissociating effects on pitch shift from those on pitch representation.

Method

Stimuli

Shepard tones were composed of nine octave-related sinusoidal components with a Gaussian spectral envelope. The spectral envelope was linear on the amplitude scale and logarithmic on the frequency scale. It was fixed for all sounds, centered at 1046.6 Hz, with a standard deviation of one \log_2 unit. This corresponds to lower and upper half-amplitude cutoff points at 463 and 2367 Hz, respectively. Each trial contained two tones, a standard and comparison. Four base frequency conditions were used for the standard, so that the F_b s were equally spaced within the octave from 65.41 to 130.82 Hz. Comparison tones were obtained by shifting each F_b in one-semitone (st) steps, to obtain all 11 possible comparison tones (the 12th would be identical to the standard). The duration of each tone was 125 ms, including 5-ms raised-cosine onset and offset ramps. The intertone silent interval within a pair was 125 ms, and the presentation level was 65 dB SPL (A-weighted).

Procedure

On each trial a tone pair was presented, and listeners reported whether the first or second tone was higher in pitch. Unbeknownst to the listener, experimental blocks were organized into sequences of ten trials. The F_b of the standard was fixed for the duration of a sequence of trials but was counterbalanced randomly across sequences. The sequence types, which will from now on be termed “conditions,” are illustrated in Fig. 2. In the *fixed* condition, tone pairs with an interval of 6 st were presented. In the *random* condition, intervals from 1 to 11 st were presented in random order. In the *increasing* condition, intervals from 1 to 11 st were presented in an ordered manner, with 1-st increases between successive trials. In the *decreasing* condition, intervals from 11 to 1 st were presented in an ordered manner, with 1-st decreases between successive trials. Importantly, in all conditions, the presentation order of the standard and comparison tones within each pair was random. Finally, in all conditions, one interval was omitted at random (counterbalanced across sequences within a condition), so that ten trials were presented per sequence. Responses were self-paced, with a delay of

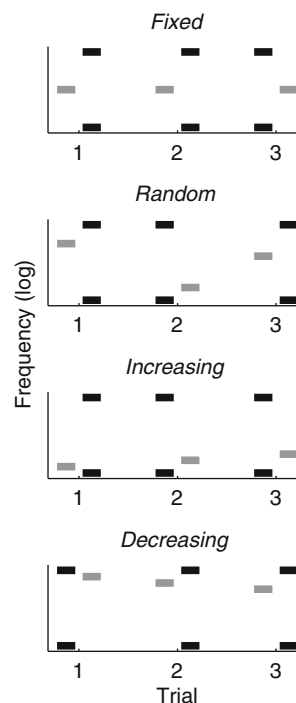


Fig. 2 Schematics of the experimental conditions. For each sequence type, the first three trials (out of ten) are illustrated. Due to the cyclic nature of the stimulus across the frequency axis, spectrograms are restricted to one octave. In all conditions, sequences of Shepard tone pairs were presented in which the interval between the standard tone (black) and comparison tone (gray) was manipulated: Fixed, the interval was fixed at 6 semitones (st); Random, intervals were presented in random order; Increasing, the interval was increased in 1-st steps; Decreasing, the interval was decreased in 1-st steps. The presentation order of the standard and the comparison tones within a pair was random

250 ms between a response and the next trial. The order of conditions within a block was randomized, and no indication was given that a sequence had started or ended. No feedback was provided.

Each interval was repeated 40 times in the random and ordered conditions, and 440 times in the fixed condition, resulting in 1,760 trials in total (44 sequences apiece for the fixed, random, and each of the ordered conditions). The experiment was divided into eight blocks of 220 trials.

Screening test

Before taking part in the experiment, listeners were tested on their ability to report pitch shifts in pairs of unambiguous Shepard tones and pure tones, since some variability was expected on this task (Semal & Demany, 2006; Shepard, 1964). In one condition, Shepard tones with random F_b in a one-octave range above 65.41 Hz were used. In another condition, pure tones with a random frequency in a one-octave range above 1046.6 Hz were used, corresponding to the region of high spectral amplitude in the main experiment.

Intervals of 1, 2, and 3 st were presented in random order, with ten repeats per interval and stimulus type. Participants with 80% or more correct for the 1-st condition, in both the Shepard and pure-tone conditions, were recruited for the main experiment.

Participants

Fourteen self-reported normal-hearing listeners with a mean age of 25.43 years ($SE = 0.40$) participated in the experiment. Four of the participants were excluded because they did not pass the screening test. Seven out of the remaining ten participants had not previously taken part in experiments involving Shepard tones, and three had taken part in an unreported pilot. The experiment was conducted according to the guidelines of the Declaration of Helsinki.

Apparatus

Listeners were tested individually in a double-walled sound-insulated booth (Industrial Acoustics Company). The stimuli were delivered diotically through an RME Fireface 800 sound card, at 16-bit resolution and a 44.1-kHz sample rate, through Sennheiser HD 250 Linear II headphones. Sound level was calibrated with a Brüel & Kjær (2250) sound level meter and a Brüel & Kjær ear simulator (4153).

Data analysis

When appropriate, psychometric functions were fitted to the raw data. Weibull functions were used to model “percentage choice”, and fits were obtained using the `psignifit` toolbox for MATLAB (see <http://bootstrap-software.org/psignifit/>), which implements the maximum-likelihood method described by Wichmann and Hill (2001). The upper bound ($1 - \lambda$) and lower bound (γ) of the Weibull function were included as parameters. These were constrained to lie between 0 and .05, and an initial value of .01 was used. The goodness-of-fit test suggested by Wichmann and Hill was also applied. From each fitted function, 999 Monte Carlo data sets were generated and then used to estimate the expected distribution of a deviance statistic for that function (Wichmann & Hill, 2001). A case was rejected if the cumulative probability estimate of the deviance exceeded .99.

Results

Fixed condition

In the fixed condition, the interval between the pair of tones on a trial was fixed at 6 st, but the order of the standard and comparison within a trial was random. We first computed the

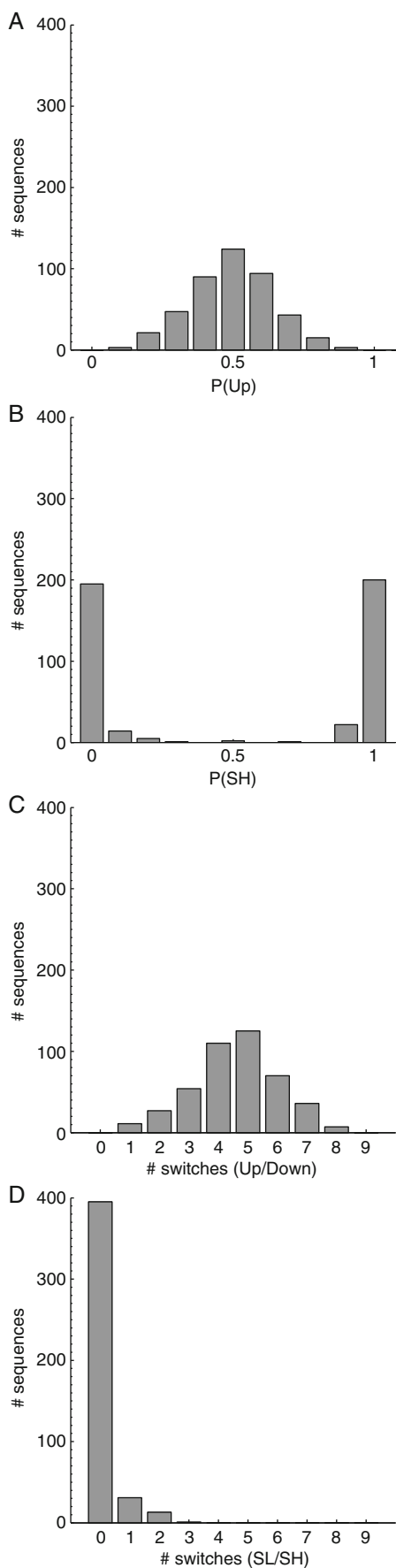
proportion of “up” responses (2nd tone higher) provided by listeners for a sequence, noted as $P(\text{Up})$. Figure 3A displays the histogram of $P(\text{Up})$ values compiled for all listeners and sequences of the fixed condition. Responses were randomly distributed around .5, indicating that listeners gave as many “up” as “down” responses.

Next, we recoded the “up” and “down” responses into “standard higher” and “standard lower” responses. The proportion of “standard higher” responses, noted as $P(\text{SH})$, was computed for each sequence. Figure 3B displays the histogram of $P(\text{SH})$ values compiled for all listeners and sequences. A clear bimodal distribution is observed, with most sequences being associated with a $P(\text{SH})$ close to either 0 or 1. Therefore, in the vast majority of cases, the perception of the standard relative to the comparison was maintained throughout the sequence, irrespective of the standard and comparison tones’ presentation order.

Another illustration of the same finding is provided in Figs. 3C and D. First, we computed the number of times that the responses switched from up to down, or the reverse. Figure 3C displays the histogram of the number of switches per sequence, compiled for all listeners and sequences. A random distribution was observed, centered around five switches per sequence. Then we recoded the responses and computed the number of times per sequence that the report switched from standard higher to standard lower, or the reverse. Figure 3D displays the histogram of the numbers of switches per sequence, compiled for all listeners and sequences of the fixed condition. Within sequences of ten trials, switches were a rare occurrence.

A lack of switches may indicate that some listeners always responded “standard lower” or “standard higher” when presented with certain stimuli, perhaps due to idiosyncratic biases. Moreover, such idiosyncratic biases may depend on the base frequencies of the tones (i.e., their pitch class; see, e.g., Deutsch, 2013). Thus, we computed the overall bias in reporting one of the stimuli as higher, per listener and per frequency condition. Figure 4 shows the result of this analysis. As expected (Deutsch, 1987, 2013), a broad range of idiosyncratic biases was found. Some listeners were strongly biased to always hear some stimuli as higher (values close to 0 or 1 in Fig. 4), but there were also several instances without any clear idiosyncratic bias. For such cases, listeners maintained a “standard higher” or “standard lower” response within a sequence (Figs. 3B and D), but varied this response across sequences (Fig. 4).

In summary, perception was highly stable within a sequence when it was measured in terms of “standard higher” or “standard lower” (Figs. 3B and D), but followed a random pattern when measured as “up” and “down” responses (Figs. 3A and C). This strongly suggests that the distribution in Fig. 3A simply reflected the random ordering of the standard and comparison tones within a trial. Thus, from now on,



◀ **Fig. 3** Results for the fixed condition. (A) Proportions of “up” responses, $P(\text{Up})$, computed for each sequence and compiled across listeners. (B) Proportions of “standard higher” responses, $P(\text{SH})$, computed for each sequence and compiled across listeners. (C) Numbers of switches between “up” and “down” responses, computed for each sequence and compiled across listeners. (D) Numbers of switches between “standard lower” and “standard higher” responses, computed for each sequence and compiled across listeners

data will be described in terms of the proportion of “standard higher” responses, $P(\text{SH})$.

Random, increasing, and decreasing conditions

The average $P(\text{SH})$ was computed for each interval, per listener and per condition (ten listeners and three conditions). The $P(\text{SH})$ curves as a function of interval were fitted with Weibull psychometric functions (Wichmann & Hill, 2001; see the Method section). A goodness-of-fit test (see the Method section) established a lack of fit for five out of the 30 cases. Visual inspection revealed that all of the rejected cases (and only the rejected cases) displayed a nonmonotonicity at the extremes of the $P(\text{SH})$ curves, corresponding to small intervals between the tones. This reflects lesser accuracy for some listeners in judging small intervals, even though the intervals are nonambiguous. Such an observation is fully consistent with the original report of Shepard (1964). These cases were excluded from the analyses requiring curve fitting. However, note that Fig. 6 below presents the unfitted data without exclusions. From the remaining cases, the interval for which the Shepard tone pairs were at their most ambiguous was computed from the fitted function. This corresponds to the

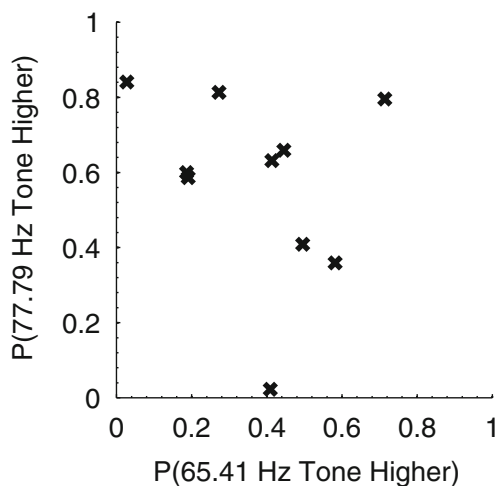


Fig. 4 Idiosyncratic biases in the fixed condition. The data from the fixed condition were split in terms of base frequency, F_b , and listener. Four base frequencies were used for the standard tone, but in analysis these collapse to two cases, since the F_b s at a half-octave distance simply exchange the roles of standard and comparison. The bias measure plotted here was computed as the proportion of trials on which the base frequency was heard as being higher in pitch, for individual listeners (crosses). Values deviating from .5 indicate the presence of an idiosyncratic bias

estimated interval for which $P(\text{SH}) = .5$ —that is, the point of subjective indifference, noted as the *PSI*.

These results are presented in Fig. 5. In the random condition (Fig. 5A, middle curve), intervals between 1 and 11 st were presented in random order. Listeners reported more often the smaller of the two possible frequency shifts between the components of Shepard tones. When the interval was small, listeners reported that the standard tone was lower; when it was large, the standard was reported as higher. For more ambiguous intervals, either response occurred.

In the ordered sequences, the interval between standard and comparison was increased from 1 to 11 st (increasing; see Fig. 5A, rightward-facing arrow) or decreased from 11 to 1 st (decreasing; leftward-facing arrow in Fig. 5A). The responses to the starting intervals of all sequences (1 st for increasing and 11 st for decreasing) were strongly biased, in that listeners reported the smaller of the two possible shifts, just as in the random condition. This initial bias persisted for subsequent percepts. Notably, the initial bias almost completely determined perception for the fully ambiguous interval at 6 st. The bias persisted even for intervals that favored the opposite percept in the random condition.

To further quantify the influence of context on the same stimulus, we examined the unfitted $P(\text{SH})$ at the 6-st interval, with all listeners' data included. This revealed a large difference between the increasing and decreasing conditions: The mean $P(\text{SH})$ was .03 ($SE = .01$) in the increasing condition, but .95 ($SE = .02$) in the decreasing condition.

Finally, we quantified hysteresis using an independent-groups analysis of variance (ANOVA) on the *PSI* (Fig. 5B), with condition (random, increasing, or decreasing) as the independent variable. The *PSI* was used for the ANOVA because individual $P(\text{SH})$ values would not be independent from each other in the case of hysteresis. The independent-groups analysis was necessary to take into account exclusions. The main effect of condition on *PSI* was highly significant, with a large effect size [$F(2, 22) = 157.97, p < .001, \eta^2 = .93$].

Omissions

In all sequences, one interval was omitted at random (balanced across sequences and conditions). If hysteresis was due to a response strategy whereby listeners waited a certain number of trials before switching from “standard lower” to “standard higher,” or vice versa, a shift of the psychometric curve should be observed, depending on which interval was omitted.

Figure 6 illustrates the effect of omissions. The $P(\text{SH})$ was computed for each listener, but this time splitting the data in two: that is, into sequences in which an interval below 6 st was omitted, and sequences in which an interval greater than 6 st was omitted. Sequences in which the interval at 6 st was omitted were excluded from the analysis.

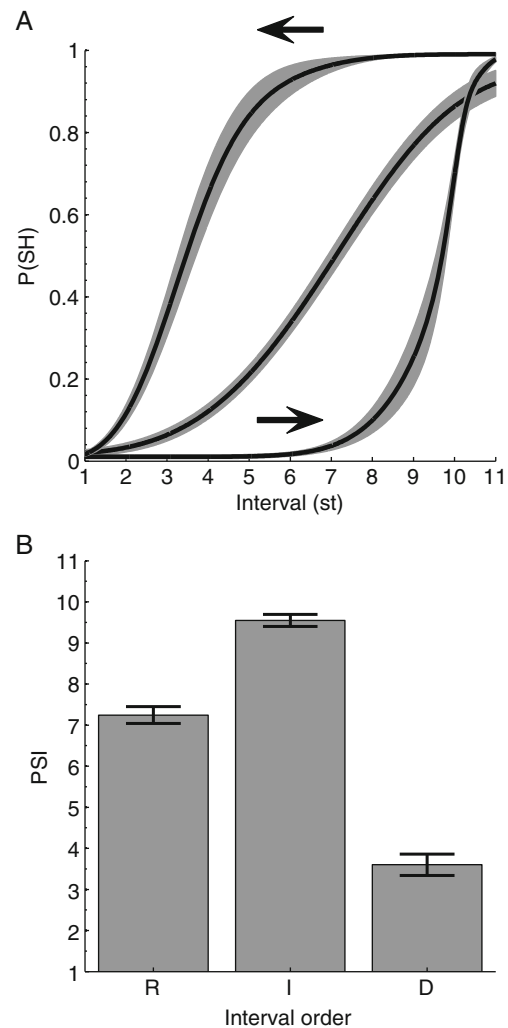


Fig. 5 Results for the random, increasing, and decreasing conditions. (A) $P(\text{SH})$ as a function of the interval between the standard and comparison tones. Fitted curves are shown for the random (middle curve), increasing (rightward-facing arrow), and decreasing (leftward-facing arrow) conditions, with the shaded areas displaying the standard errors of the means. (B) Points of subjective indifference (*PSI*s) for the random (R), increasing (I), and decreasing (D) conditions

It is clear from Fig. 6 that omissions had no sizeable effect. Nevertheless, we quantified the effect of the omission on the *PSI*. We split the data on the basis of the condition (increasing or decreasing) and where the omission occurred (before or after 6 st) and computed the *PSI* for each case. Thirteen out of the 40 cases (ten listeners, two conditions, and two omissions) were excluded due to a lack of fit. This is about double the number of rejected cases in the main analysis, consistent with the fact that we split the data in two. Again, exclusions corresponded to nonmonotonicity in the $P(\text{SH})$ curves.

An independent-groups ANOVA was performed with omission (before or after 6 st) and condition (increasing or decreasing) as independent variables. As predicted, we found a significant main effect of condition [$F(1, 23) = 496.28, p < .001$], no significant main effect of omission [$F(1, 23) = 0.39,$

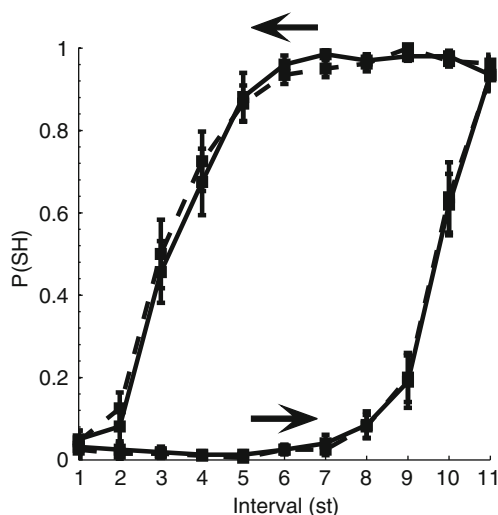


Fig. 6 Effect of omissions: *P*(SH) values averaged across all listeners are shown as a function of interval for the increasing (rightward-pointing arrow) and decreasing (leftward-pointing arrow) conditions, for sequences in which an interval below 6 st was excluded (dotted lines) and sequences in which an interval above 6 st was excluded (solid lines)

$p = .54$], and no significant interaction between condition and omission [$F(1, 23) = 0.62, p = .44$]. This indicates that the psychometric functions remained stable, irrespective of omitted interval. The PSI was not reached after a fixed number of responses, but rather at a specific interval.

Molecular analysis of the random condition

The random condition can be viewed as a baseline for the ordered conditions. However, it can also be used to assess whether previous trials tended to influence current responses, since each trial was preceded by a random interval.

We assessed across-trial effects using a molecular analysis (Dittrich & Oberfeld, 2009). A binary logistic regression was performed on each individual data set, in order to predict the response on the current trial. Two models were investigated. Model 1 included the interval on the current trial and the percept (standard lower or higher) on the four most recent trials as predictors. Model 2 included the interval on the current trial and the intervals (in semitones) on the four most recent trials as predictors. Separate regression analyses were conducted for each participant. Each model was assessed for each participant using the Hosmer–Lemeshow goodness-of-fit test (Dittrich & Oberfeld, 2009). Briefly, observations were divided into ten evenly sized bins on the basis of their probability, predicted from the regression model. The hypothesis that the observed number of events in each group was different from the expected number of events was tested with a chi-square test. Significant differences ($p < .05$) were taken as a poor fit of the model to the data. Model 1 led to a lack of fit for two out of ten participants, whereas Model 2 produced acceptable fits in all cases. Thus, Model 2 was selected.

The weights from the regression model, normalized so that their sum equals one, are provided in Table 1. Positive weights were observed for all predictors. Their statistical reliability was estimated using one-sample Bonferroni-corrected *t* tests. Significance was accepted at a *p* value of .05/5. The outcomes of these tests are also listed in Table 1. Weights were significantly different from zero for the interval of the current trial (I) and the two most recent intervals (Intervals-1 and Interval-2), becoming nonsignificant for the third most recent interval (Interval-3).

Note that, in this analysis, the weights obtained for the current interval and those for the previous intervals are not necessarily comparable. The interval on the current trial is likely to influence the response in a monotonic way, mirroring the results of the random condition (Fig. 5A). However, there is no strong reason to assume that this would be the case for past intervals: The smallest or largest past intervals might not be those inducing the most potent biases. In any case, the comparison between past and present weights is not crucial, since the main aim of the analysis was to assess a potential influence of previous trials on the current response. The outcome of this aspect of the analysis is clear: The response on the current trial was influenced by the two most recent trials, with decreasing effectiveness for more distant trials.

Discussion

A new method to measure auditory hysteresis in pitch judgment has been reported. This method aimed to address the pitfalls identified in previous investigations, in order to isolate perceptual hysteresis from responses bias (Hock et al., 1993). Shepard tone pairs were used, with a simple parametric manipulation to control the degree of ambiguity in each trial: the interval between the standard and comparison tone. This manipulation was obfuscated by the complex structure of the experimental blocks: random order between the standard and comparison on each trial; four randomly interleaved conditions; and omissions within sequences. In particular, due to the random ordering of the standard and comparison, listeners were forced to vary between “up” and “down” responses even for less ambiguous trials. Importantly, this resulted in a pattern

Table 1 *t* tests for predictors in a molecular regression analysis

Predictor	Mean (SE)	<i>t</i> (<i>df</i> = 9)
Interval	65 (.03)	17.74*
Interval-1	14 (.01)	9.96*
Interval-2	08 (.01)	5.12*
Interval-3	07 (.02)	3.00
Interval-4	03 (.02)	1.36

* $p < 0.01$

of result that cannot be attributed to hysteresis in the pattern of keypresses. If listeners reached a region of uncertainty when the stimulus entered the ambiguous range, a strategy that consisted of relying on the previous keypresses would have resulted in no recorded hysteresis, unlike in previous investigations. Another type of bias, common to most hysteresis experiments, could still be considered. If listeners were able to detect the reoccurrence of the standard tone in a sequence of trials, and if they chose to assign a label to the standard according to their response in the first trial, then a decisional bias might occur whereby the chosen label would be maintained throughout the ambiguous region. This interpretation cannot be ruled out for the fixed and ordered sequences. However, it cannot account for hysteresis in the random sequences, since the molecular analysis showed a consistent influence of the two most recent trials on current responses, even though the label of the standard would have changed randomly from trial to trial.

In summary, strong hysteresis was observed in spite of the variable response patterns in all experimental conditions, ruling out response biases related to keypresses. Randomization of the physical parameter of interest in one condition cast doubts on an interpretation based on decisional biases. We thus suggest that the hysteresis was perceptual in nature, a conclusion consistent with all of the observed data. Our results were also consistent with the reports of Giangrande et al. (2003), in the sense that context effects were assimilative. A given percept was maintained in the face of ambiguity, and even of conflicting sensory evidence.

In addition to removing potential confounds, we extended the findings of Giangrande et al. (2003) on several counts. First, the hysteresis observed here was stronger, perhaps because of differences in the stimulus parameters (spectral shape, duration of each tone, intertone interval, and randomization of the standard tone frequency) or due to our pool of listeners. It is notable that our pool of listeners did contain individuals with strong idiosyncratic biases for some of the stimuli (see results for the fixed condition in Fig. 4; cf. Deutsch, 2013). Nevertheless, even for those individuals, hysteresis fully dominated over the idiosyncratic biases. Second, and perhaps more interestingly, our method specified what the context actually modulated. Hysteresis was not related to the perception of upward versus downward shifts. Our analysis of the fixed condition showed that the direction of shift heard on one trial did not have an effect on the direction heard on the next trial; rather, the pitch of the standard relative to the comparison was biased. Such a bias was observed irrespective of the standard–comparison or comparison–standard pattern presented to the listener, and hence of the “up” or “down” reported percept. This finding explains the lack of context effect observed in some other studies using Shepard tones. Dawe et al. (1998) and Repp and Thompson (2010) attempted to bias the direction of pitch shift, either by adapting

one direction (Dawe et al., 1998) or by presenting unambiguous pitch shifts before an ambiguous shift (Repp & Thompson, 2010). Third, we quantified the influence of past trials on the current percept using a molecular analysis. These results show that the bias was not simply passed on from trial to trial, but that perception depends on sensory history from several recent trials. The underlying effect therefore constitutes a memory-like phenomenon, which accumulates over time and is not entirely eliminated by intervening stimuli with opposing biases.

The root cause of the hysteresis effect remains unclear. In an experiment in which the direction of pitch shift was measured in pairs of pure tones around threshold, Raviv, Ahissar, and Loewenstein (2012) observed an influence of previous trials on current judgments. They accounted for the effect with a regression-to-the-mean explanation (Hollingworth, 1910), whereby the frequency representation of the current tone was attracted to the running average of the frequency representations of previous tones. Qualitatively, this may be consistent with the present observations: Past intervals may have “attracted” the interval presented on the current trial, if each frequency component of the comparison tone regressed toward the running mean of the closest component in previous trials. This would induce assimilative hysteresis. A different account could involve auditory streaming (Moore & Gockel, 2012): Because of frequency proximity, it is plausible that some form of perceptual binding was established between neighboring tone components during the early trials of a hysteresis sequence. Such a binding could have persisted throughout the ambiguous parts of the sequence, because of the gradual nature of the shift in frequency, hence causing hysteresis. Both hypotheses remain to be further specified, however, and tested experimentally.

The remarkably large hysteresis effect observed here suggests that adaptive processes may play an important role in the ongoing perception of sound, even above threshold. The method presented here could provide a useful tool for investigating the neurophysiological bases of such processes, since it enables the experimenter to induce large pitch changes in the perception of the same stimulus.

Author note We extend our thanks to Daniel Oberfeld, who suggested the idea of a molecular analysis and provided scripts to perform the computations. This work was funded by a grant from Entendre SAS to the first author, and by ERC Grant Nos. ADAM 295603, ANR-10-LABX-0087 IEC, and ANR-10-IDEX-0001-02 PSL*. The authors declare no conflict of interest.

References

- Dawe, L., Platt, J. R., & Welsh, E. (1998). Spectral-motion aftereffects and the tritone paradox among Canadian subjects. *Perception & Psychophysics*, *60*, 209–220.

- Deutsch, D. (1987). The tritone paradox: Effects of spectral variables. *Perception & Psychophysics*, *41*, 563–575.
- Deutsch, D. (2013). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 349–411). San Diego: Elsevier.
- Deutsch, D., Moore, F. R., & Dolson, M. (1986). The perceived height of octave-related complexes. *The Journal of the Acoustical Society of America*, *80*, 1346–1353.
- Dittrich, K., & Oberfeld, D. (2009). A comparison of the temporal weighting of annoyance and loudness. *Journal of the Acoustical Society of America*, *126*, 3168–3178.
- Giangrande, J., Tuller, B., & Kelso, J. (2003). Perceptual dynamics of circular pitch. *Music Perception*, *20*, 241–262.
- Hock, H. S., Kelso, J. S., & Schöner, G. (1993). Bistability and hysteresis in the organization of apparent motion patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 63–80. doi:10.1037/0096-1523.19.1.63
- Hollingworth, H. (1910). The central tendency of judgment. *Journal of Philosophy, Psychology and Scientific Method*, *7*, 461–469.
- Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, *16*, 305–312.
- Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *Journal of the Acoustical Society of America*, *120*, 2801–2817.
- Huang, J., & Holt, L. L. (2012). Listening for the norm: Adaptive coding in speech categorization. *Frontiers in Psychology*, *3*(10), 1–6. doi:10.3389/fpsyg.2012.00010
- Kanai, R., & Verstraten, F. (2005). Perceptual manifestations of fast neural plasticity: Motion priming, rapid motion aftereffect and perceptual sensitization. *Vision Research*, *45*, 3109–3116.
- Laing, E. J. C., Liu, R., Lotto, A. J., & Holt, L. L. (2012). Tuned with a tune: Talker normalization via general auditory processes. *Frontiers in Psychology*, *3*(203), 1–9. doi:10.3389/fpsyg.2012.00203
- Leopold, D., & Logothetis, N. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences*, *3*, 254–264.
- Leopold, D. A., Wilke, M., Maier, A., & Logothetis, N. K. (2002). Stable perception of visually ambiguous patterns. *Nature Neuroscience*, *5*, 605–609.
- Maloney, L. T., Dal Martello, M. F., Sahn, C., & Spillmann, L. (2005). Past trials influence perception of ambiguous motion quartets through pattern completion. *Proceedings of the National Academy of Sciences*, *102*, 3164–3169. doi:10.1073/pnas.0407157102
- Moore, B. C. J., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B*, *367*, 919–931.
- Noest, A. J., van Ee, R., Nijs, M. M., & van Wezel, R. J. (2007). Percept-choice sequences driven by interrupted ambiguous stimuli: A low-level neural model. *Journal of Vision*, *7*(8), 1–14. doi:10.1167/7.8.10
- Ragozzine, F., & Deutsch, D. (1994). A regional difference in perception of the tritone paradox within the United States. *Music Perception*, *12*, 213–225.
- Raviv, O., Ahissar, M., & Loewenstein, Y. (2012). How recent history affects perception: The normative approach and its heuristic approximation. *PLoS Computational Biology*, *8*, e1002731. doi:10.1371/journal.pcbi.1002731
- Repp, B. H. (1997). Spectral envelope and context effects in the tritone paradox. *Perception*, *26*, 645–665.
- Repp, B. H., & Thompson, J. M. (2010). Context sensitivity and invariance in perception of octave-ambiguous tones. *Psychological Research*, *74*, 437–456.
- Schwartz, J.-L., Grimault, N., Hupé, J.-M., Moore, B. C. J., & Pressnitzer, D. (2012). Multistability in perception: Binding sensory modalities, an overview. *Philosophical Transactions of the Royal Society B*, *367*, 896–905. doi:10.1098/rstb.2011.0254
- Semal, C., & Demany, L. (2006). Individual differences in the sensitivity to pitch direction. *Journal of the Acoustical Society of America*, *120*, 3907–3915.
- Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, *36*, 2346–2353.
- Snyder, J. S., Carter, O. L., Hannon, E. E., & Alain, C. (2009). Adaptation reveals multiple levels of representation in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1232–1244. doi:10.1037/a0012741
- Snyder, J. S., Carter, O. L., Lee, S.-K., Hannon, E. E., & Alain, C. (2008). Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1007–1016. doi:10.1037/0096-1523.34.4.1007
- Wichmann, F., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*, 1293–1313. doi:10.3758/BF03194544