Chapter 18
Putting the Tritone Paradox into Context: Insights from Neural Population Decoding and Human Psychophysics

Bernhard Englitz, S. Akram, S.V. David, C. Chambers, Daniel Pressnitzer, D. Depireux, J.B. Fritz, and Shihab A. Shamma

Abstract The context in which a stimulus occurs can influence its perception. We study contextual effects in audition using the tritone paradox, where a pair of complex (Shepard) tones separated by half an octave can be perceived as ascending or descending. While ambiguous in isolation, they are heard with a clear upward or downward change in pitch, when preceded by spectrally matched biasing sequences. We presented these *biased Shepard pairs* to awake ferrets and obtained neuronal responses from primary auditory cortex. Using dimensionality reduction from the neural population response, we decode the perceived pitch for each tone. The bias sequence is found to reliably shift the perceived pitch of the tones away from its central frequency. Using human psychophysics, we provide evidence that this shift in pitch is present in active human perception as well. These results are incompatible with the standard absolute distance decoder for Shepard tones, which would have predicted the bias to attract the tones. We propose a relative decoder that takes the stimulus history into account and is consistent with the present and other data sets.

B. Englitz, PhD (✉) • S. Akram • D. Depireux • J.B. Fritz
Institute for Systems Research, University of Maryland,
A.V. Williams Bldg., College Park, MD 20742, USA
e-mail: benglitz@gmail.com

S.V. David
Institute for Systems Research, University of Maryland,
A.V. Williams Bldg., College Park, MD 20742, USA

Oregon Hearing Research Center,
Oregon Health & Science University, Portland, OR, USA

C. Chambers • D. Pressnitzer
Institute for Systems Research, Equipe Audition, Ecole Normale Superieure, Paris, France

S.A. Shamma
Department of Electrical and Computer Engineering,
Institute for Systems Research, University of Maryland, College Park, MD, USA

B.C.J. Moore et al. (eds.), *Basic Aspects of Hearing*,
Advances in Experimental Medicine and Biology,
DOI 10.1007/978-1-4614-1590-9_18,

1 Introduction

In the present study, we address the physiological basis of the auditory ‘tritone paradox’ (Deutsch et al. 1986), in which a sequence of two tonal stimuli can be perceived as ascending or descending in pitch. Various factors, including linguistic background (Deutsch 1991, 1994), amplitude envelope (Repp 1997), and stimulus context (Chambers and Pressnitzer 2011), have been shown to influence the perceived direction. The present focus is the influence of stimulus history on the directionality percept. We address this question by preceding the ambiguous pair with a sequence of tonal stimuli (‘bias’) in a limited spectral range, which has been demonstrated to reliably influence human perception (Chambers and Pressnitzer 2011). Identical stimuli are presented to ferrets in physiological experiments and humans in psychophysical experiments.

We find the bias to influence the perceived pitch of the tones in the ambiguous pair in a repulsive manner, i.e. the pitch of the tones shifts away from the bias. These results from neuronal population decoding are supported by human psychophysics.

These results allow us to evaluate different decoders for directionality in pitch. The most intuitive decoder, which relies on the minimal—ascending or descending—distance between the two tonal stimuli, is not consistent with the bias-induced increase in distance on the biased side of the pitch circle. We propose a more general decoder, which takes the stimulus history into account and allows for a transitive conjunction of two relative directionality judgements to provide an explanation for the central findings on the tritone paradox and its context dependence.

2 Methods

2.1 Physiology and Psychophysics: Stimulation and Recording

Both human and animal experiments were conducted according to the ethical guidelines of the University of Maryland. Stimuli were presented in sound-attenuated chambers over calibrated speakers. Extracellular neuronal recordings were collected from seven awake ferrets using chronically implanted multielectrode arrays. In human experiments, an optimized sampling technique was applied to rapidly estimate the psychophysical curve with high precision (Poppe et al. 2012) (Fig. 18.1).

2.2 Acoustic Stimuli

All stimuli were composed of sequences of Shepard tones (Shepard 1964). A Shepard tone is a complex tone built as the sum of octave-spaced pure tones (a flat spectral envelope was used here). A Shepard tone can be characterized by its position

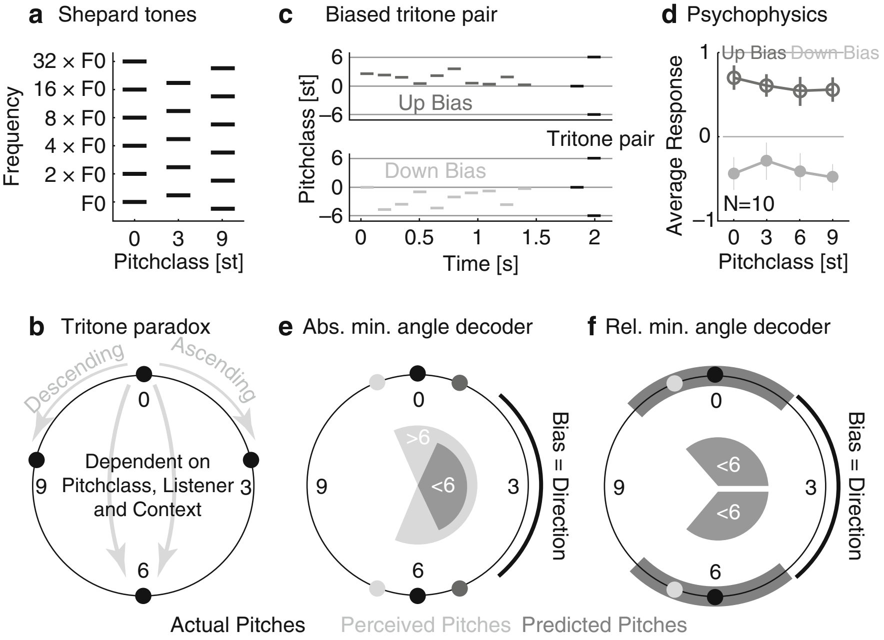

Fig. 18.1 The tritone paradox and contextual modulation of the perceived change in pitch. (**a**) The ambiguity in the tritone paradox stems from the Shepard tones' repetitive structure (see Sect. 2). (**b**) Steps of up to almost 6 st are perceived as up/down steps, whereas the half-octave (tritone) step is ambiguous and influenced by various factors (acoustic context, pitch class, etc.). (**c**) Preceding an ambiguous Shepard pair by a sequence of Shepard tones within the half octave above ('up bias', *dark grey*) or below ('down bias', *light grey*), the first tone (**d**) influences the percept of directionality for human listeners. (**e**) As indicated in the Sect. 1 and detailed in the Sect. 4," the standard decoder assumes the absolute minimal angle between the Shepard pair to determine the perceived direction of pitch change (perceived = *light grey*, prediction = *dark grey*). We show that the perceived pitches move away (i.e. >6 st) from the bias (actual = *black*), thus necessitating a different decoder. (**f**) A relative decoder takes the acoustic context into account and makes relative judgements from the acoustic history, consistent with the present data set. The *dark grey* areas indicate acceptable perceived pitches of the tones to still be heard according to the bias for the relative decoder

in an octave, termed pitch class (in units of semitones), w.r.t. a base tone. Across the entire set of experiments the duration of the Shepard tones was 0.1 s and the amplitude 70 dB SPL.

The *biased Shepard pairs* consisted of a bias sequence ('bias') followed by an ambiguous, i.e. 6 st separated, Shepard pair. The bias precedes the pair at various temporal separations ([0.05,0.2,0.5,1]s) and consists of a sequence of Shepard tones (lengths, 5 and 10 stimuli), which are within 5 semitones above or below the first Shepard tone in the pair. These biases are called 'up' and 'down' bias, respectively, as they bias the perception of the ambiguous pair to be 'ascending' or 'descending', respectively, in pitch (Chambers and Pressnitzer 2011). Altogether we presented 32 conditions (4 base pitch classes ([0,3,6,9]st), 2 randomization, 2 bias lengths ([5,10]

stimuli), ‘up’/‘down bias) and different bias sequences, which in total contained 240 distinct Shepard tones, finely covering one octave. In the present study, we use one of the simpler versions of this contextual influence—more detailed psychophysics will be described in a forthcoming study by CC, SAS and DP.

Further, we used pitch comparison sequences in the psychophysical studies. The pitch comparison sequences consisted of a bias, a reference Shepard tone, and a target Shepard tone. The bias that preceded the reference was followed by a target (drawn from the set [−3,−2.9,…,2.9,3]st) 3 s later. Subjects were asked to report whether the target’s pitch was higher or lower than the reference’s.

2.3 Population Decoding

The perceived stimuli in the ambiguous pair were estimated from the neural responses by training a decoder on the biasing sequences and then applying the decoder to the neural response of the pair. We first build a matrix of responses which had the (240) different Shepard tones occurring in the bias running along one dimension and the neurons along the other dimension. The PCA (Principle Component Analysis) decoder performed a linear dimensionality reduction, interpreting the stimuli as examples and the neurons as dimensions of the representation. The data was projected to the first three dimensions, which represented the pitch class as well as the position in the sequence of stimuli. To assign a pitch class to the decoded stimuli of the test pair, we projected them onto the ‘pitch circle’ formed by the decoded stimuli from the bias sequences. More precisely we estimated a smoothed trajectory through the set of bias tones which was assigned a pitch class at every point, by averaging the pitch classes of the closest 10 bias stimuli weighted by their distance to the point. The pitch class of the test tone was set to the pitch class of the closest point on the trajectory.

2.4 Statistical Analysis

Nonparametric tests were used throughout the study to avoid assumptions regarding distributional shape.

3 Results

We obtained single-unit recordings from 555 neurons in the primary auditory cortex of seven awake ferrets and conducted psychophysical experiments with ten subjects under various stimulus conditions.

3.1 Neurons in Auditory Cortex Exhibit Tuning to Shepard Tones

A considerable subset of neurons in auditory cortex responded to the presentation of Shepard tones with a significant change in response rate compared to spontaneous rate (55, 43 % increased, 12 % decreased; $p < 0.05$), while 41 % of the neurons also had a significantly tuned response. A well-tuned unit is shown in Fig. 18.2a, where the firing rate varies as a function of the pitch class of the Shepard tone. Neurons typically exhibited a single peak of varying width, although multipeaked tuning curves existed as well (~30 % out of the tuned cells). Overall, the median tuning width was 2.06 [25 %, 0.82; 75 %, 6.44]st (2 SD of a Gaussian fit to the tuning). Neurons exhibited strongest tuning in onset, sustained, and offset responses in similar proportions (onset, 38 %; sustained, 33 %; offset, 29 %).

Many cells (39 %) showed significant changes in response strength over the sequence of presented stimuli (Friedman test, Fig. 18.2f), where 33 % showed a decrease (rel. to 1 st stimulus, median 31 %, range 14–45 %) over time and 6 % an increase (median 11 %, range 5–31 %). Suppression/facilitation was assessed by fitting a single time-constant exponential decay to the average PSTH of all biasing stimuli.

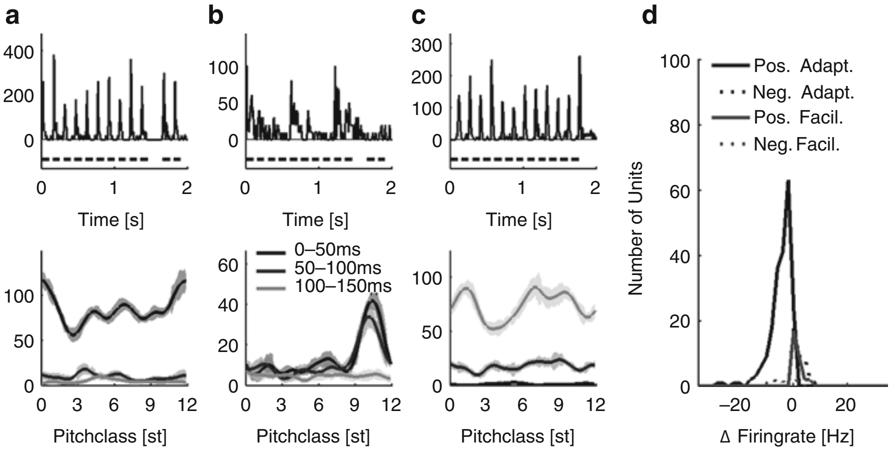

Fig. 18.2 Response of auditory cortex neurons to Shepard tones is typically tuned and adapting. Three example cells are shown in (**a–c**), an onset tuned cell (**a**), a sustained tuned cell (**b**) and an offset tuned cell (**c**). *Top panels* show an average over 10 trials to one bias sequence and the ambiguous pair after a pause of 0.2 s. Bottom panels show the responses as a function of all Shepard tones in the bias sequences (smoothed with a Gaussian fit, 0.5 st), with different grey values indicating the tuning at different times of the response. (**d**) The large majority of significantly responding cells adapted their response over the duration of the bias (*black*), while only a small minority facilitated to a smaller degree (*grey*). Positive and negative deviations from spontaneous rate are considered separately here (see text)

3.2 Bias Stimulus Repels the Tritone Pair in Pitch

The pitch class of a stimulus is encoded in the population response, via the various response properties of different neurons. We used a dimensionality reduction-based decoder to transition from the high-dimensional space of neural responses to only two dimensions. Stimuli of similar pitch class (Fig. 18.3a, different colours) occupy close regions in space. Globally, the stimuli form a slightly distorted torus, along which the pitch class changes in an orderly fashion. This circular, closed shape is expected, given that the Shepard tones form a circular stimulus space themselves.

The represented pitch class of the tones in the tritone pair generally falls into their respective regions on the pitch circle; however, they are shifted in different directions along the circle depending on which bias preceded them (Fig. 18.3b). If the bias is above the first tone in the pair, the first tone is shifted down, while the second tone shifted up, and conversely for a bias below the first tone. More simply put, the tones in the ambiguous pair shift away in pitch class from the bias sequence (127/128 comparisons between the positions of the tones across all conditions (see Sect. 2); Wilcoxon test, $p < 10–22$). The bias-induced shift was 0.62 st for the first and 0.49 st for the second tone and returned to baseline with an exponential time constant of 0.56 s (estimated based on the recovery of the spontaneous rate after the tritone pair).

The presence of the bias hence increases the distance between the tritone pair along the side of the pitch circle, which humans judge to be the direction of pitch change.

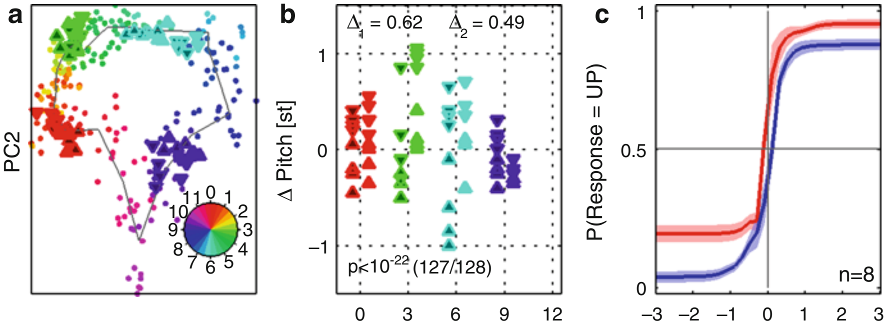

Fig. 18.3 Neural population decoding shows the bias to repel the tritone pair in pitch. (**a**) After dimensionality reduction to three dimensions, the neuronal responses to Shepard tones spanning the whole octave exhibit a circular progression of pitch classes (*colours*) in dimensions 1 and 2 (the *grey line* connects local averages every 0.5 st). (**b**) The Shepard tones in the tritone pairs (*triangles*) are mapped to their pitch classes; however, their precise position depends on the preceding bias, i.e. whether it is above (*triangle pointing up*) or below (*triangle pointing down*) the respective tone. The bias repels the tones in pitch, such that the same tone is perceptually separated by ~0.5 st between the two bias conditions (expected pitch class subtracted here to show differences). (**c**) Psychophysical results from humans exhibit a similar repulsion in perceived pitch in a pitch comparison paradigm with a 3 s delay separating reference and target (*red* = preceded by UP bias, *blue* = preceded by DOWN bias)

3.3 Human Perception in Biased Pitch Discrimination Mirrors Physiological Pitch Shifts

Since this result is unexpected, given the standard decoder, we tested whether it holds for human subjects. We modified the paradigm to test pitch perception more explicitly (rather than a local pitch change), by separating the tone pair temporally, in which case we hypothesize the bias to only influence the first—reference—tone. More precisely, subjects were asked to compare two Shepard tones which were separated by a long pause or noise (3 s) and preceded by the same bias sequences as above. The second Shepard tone was drawn from the range [−3…3]st relative to the first tone.

Pitch comparison judgements exhibited a behavior consistent with the neural decoding results. For Shepard tones preceded by *up* biases, the psychophysical curve shifted (by ~0.5 st on average) to lower pitch classes and the uncertainty for judgements on the low side increased (Fig. 18.3f, red curve, vice versa for *down* biases, blue curve). This is consistent with shifts in the same direction but variable size, which is predicted by the random nature of the bias sequence.

4 Discussion

How does an observer arrive at a judgement of whether the sequence of two Shepard tones is ascending or descending in pitch? Multiple hypotheses, i.e. decoders, are compatible with the previous psychophysical data. Using neural population analysis, we provide some physiological data to test whether the predictions of the classical decoder are compatible with human perception of the ambiguous Shepard pair in the context of a bias.

The classical decoder chooses the smaller angle between two Shepard tones on the pitch circle as the direction of pitch change, i.e. steps from 0 to up to 5.9 st are perceived as ascending, 0 to down to −5.9 st as descending, and the half-octave step remains ambiguous. For a biasing stimulus preceding a tritone pair, it predicts the corresponding angle to decrease, e.g. a 0–6 pair preceded by an 'up' bias could become a 1–5 pair and thus ascending (Fig. 18.1e).

A different hypothesis would be a relative decoder, which assesses a stimulus relative to the stimulus context (Fig. 18.1f), i.e. if the stimulus history is centred around 3 st, the ambiguous pair 0–6 is perceived as ascending, since 0 is relatively below and 6 relatively above the stimulus history, and by transitivity 6 is above 0. Hence, for a biasing stimulus preceding the tritone pair, it predicts an extended range of directional judgements in the direction of the bias. While it does not make a prediction about the absolute perceived location of the ambiguous pair, it is consistent with both decreases and increases in distance between them over a limited range. In the absence of a biasing stimulus, the first tone in the pair becomes the context for the second, and this decoder reduces to the classical one. A neural correlate of the relative decoder could be frequency-change selective cells, which could facilitate/adapt their response based on the stimulus history.

Our decoding results demonstrate that the angle between the two Shepard tones increases in the direction of the bias and are thus incompatible with the classical decoder, but consistent with the relative decoder. Further, as demonstrated by Chambers et al. (2009), the range of judgements following the bias exceeds 6 st, consistently with the relative decoder.

More generally, a repulsive effect of a spectrally limited stimulus history has been described in other contexts before, e.g. Holt (2005, *contrastive effect*). Our findings also relate to findings in the visual system (Seriès et al. 2009), where the overestimation of small differences between oriented gratings was linked to local adaptation and the 'homunculus' was also unaware of its own adaptation. Clearly, the 'Rosetta stone' to decode the 'tritone paradox' would be recordings from behaving animals, where perception and neural activity are comparable on a trial-by-trial basis.

Acknowledgment We thank Barak Shechter, Timm Lochmann, and Paul Watkins for interesting discussions.

References

Chambers C, Pressnitzer D (2011) The effect of context in the perception of an ambiguous pitch stimulus. ARO Abstract #1025

Chambers C, Park-Thompson V, Pressnitzer D (2009) Biasing perception of ambiguous pitch stimuli. BSA Short papers meeting, Liverpool, UK

Deutsch D (1991) The tritone paradox: an influence of language on music perception. Music Percept 8(4):335–347

Deutsch D (1994) The tritone paradox: some further geographical correlates. Music Percept 12(1):125–136

Deutsch D, Kuyper WL, Fisher Y (1986) The tritone paradox: its presence and form of distribution in a general population. A musical paradox, music perception, 3(3):275–280.

Holt LL (2005) Temporally nonadjacent nonlinguistic sounds affect speech categorization. Psychol Sci 16(4):305–312

Poppe S, Benner P, Elze T (2012) A predictive approach to nonparametric influence for adaptive sequential sampling of psychophysical experiments. J Math Psychol 56:179–195

Repp BH (1997) Spectral envelope and context effects in the tritone paradox. Perception 26(5):645–665

Seriès P, Stocker AA, Simoncelli EP (2009) Is the homunculus "aware" of sensory adaptation? Neural Comput 21(12):3271–3304

Shepard RN (1964) Circularity in judgments of relative pitch. J Acoust Soc Am 36(12):2346