The effect of stimulus range on two-interval frequency discrimination

William J. Matthews

and

Neil Stewart

Department of Psychology

University of Warwick

Coventry

CV4 7AL

United Kingdom

Tel: +44/0 24 765 23096

Fax: +44/0 24 765 24225

Email: w.j.matthews@warwick.ac.uk neil.stewart@warwick.ac.uk

Abstract

It has traditionally been thought that performance in two-interval frequency discrimination tasks decreases as the range over which the standard tone varies is increased. Recent empirical evidence and a re-examination of previous results suggest that this may not be the case. The present experiment found that performance was significantly better when the standard roved over a wide range (1500 Hz) than a narrow range (30 Hz). This pattern cannot readily be accommodated by traditional models of frequency discrimination based on memory or attention, but may be explicable in terms of neural plasticity and the formation of perceptual anchors.

PACS numbers: 43.66.Fe

The effect of stimulus range on two-interval frequency discrimination

1. Introduction

In two-interval frequency discrimination experiments the participant is required to discriminate a temporally ordered pair of tones (e.g., a standard tone followed by a comparison tone) which differ only in frequency. It is well established that performance in such tasks is better with a fixed standard stimulus than when the standard roves over a range of frequencies from trial to trial (e.g., Harris, 1952. Note that here, and throughout this article, we use 'roving standard' to mean that on each trial a standard was selected from among a fixed number of unchanging frequencies, spread across a particular range. This is distinct from the use of a roving standard whose frequency is a random number sampled from a uniform distribution. As we discuss below, this distinction may be very important).

It is not clear how frequency discrimination is influenced by the range over which the standard roves (see Amitay et al., 2005; Jesteadt & Bilger, 1974), although there are a number of reasons for expecting an increase in range to impair frequency discrimination. Studies of intensity perception have established that when the range over which the standard varies is increased, discrimination performance declines (Berliner & Durlach, 1973; Berliner, Durlach, & Braida, 1977; Jesteadt & Bilger, 1974). It might reasonably be expected that frequency discrimination will follow the same pattern. Furthermore, several theoretical models of frequency discrimination, some emphasizing memory processes and others emphasizing auditory attention, predict that increases in stimulus range will reduce discrimination accuracy.

Perhaps the most well-known account of discrimination performance to emphasize the importance of memory processes is the theory of intensity perception developed by Durlach and

Braida (1969). According to this model, performance depends on the use of two distinct memory modes. In the *sensory trace* mode, the participant attempts to maintain a trace of the standard tone and compares the comparison tone with this memory. In the *context coding* mode, the participant judges the comparison tone with respect to the general context of sounds in the experiment. The amount of noise in the trace mode depends upon the temporal interval between the standard and comparison tones; the noise in the context coding mode depends upon the range of stimuli presented. Thus, for a fixed temporal interval between standard and comparison tones, discrimination performance is predicted to be worse with a roving standard. Moreover, performance is predicted to decline as the range over which the standard varies is increased. Studies of two-interval intensity discrimination have provided support for this model (Berliner & Durlach, 1973; Berliner et al. 1977) and it has been argued that the model also applies to frequency discrimination (Jesteadt & Bilger, 1974).

Theories of frequency discrimination which emphasize auditory attention also predict that performance should be better when the standard roves over a narrower range of frequencies. When the participant is presented with the same standard on every trial, he or she can focus attention on a narrow frequency region. When the standard varies, the participant may either broaden his or her attentional band to cover a wider range of frequencies (e.g., Botte, 1995), or attempt to monitor more than one band, ignoring frequencies which fall in between (e.g., Macmillan & Schwartz, 1975). If the participant adopts the former strategy, performance will decrease as the range of frequencies increases because as the attentional band is broadened, resolution is diminished. If the participant instead elects to attend to two or more attentional bands simultaneously then the situation is more complicated and performance will depend on the precise placement and width of the bands, but will generally deteriorate as the number of bands and the distance between them increases.

Despite the prediction from both memory- and attention-oriented accounts of frequency perception that performance should be inversely related to stimulus range, there is relatively little evidence from two-interval discrimination tasks against which to test this prediction. In an assessment of the applicability of Durlach and Braida's (1969) theory of intensity perception to frequency discrimination, Jesteadt and Bilger (1974) examined the effect of stimulus range on frequency discrimination performance in two-interval forced choice (2IFC) and same-different tasks. In the fixed standard condition of their experiment, the same 1000 Hz tone was used as the standard on all trials; in the *jittered* condition, the standard varied over a relatively narrow range of frequencies (980, 990, 1000, 1010, 1020 Hz); in the *roving* condition, the range of standards was much wider (795, 890, 1000, 1120, and 1260 Hz). Jesteadt and Bilger reported that, as for intensity discrimination, frequency discrimination declined as the range over which the standards varied increased, and argued that Durlach and Braida's (1969) theory of intensity perception applies to frequency perception, too. Since Jesteadt and Bilger's work, it has generally been accepted that frequency discrimination is better when the standard roves over a narrow range than when it varies over a more disparate set of frequencies – perhaps partly because of the strong theoretical reasons for expecting this result. However, closer examination suggests that their data do not convincingly demonstrate worsening frequency discrimination when the range over which the standards varies is increased. Jesteadt and Bilger did not use inferential statistics to compare the performance in different conditions, and an examination of the qualitative pattern of their results reveals that, of their four participants, one showed uniformly poor performance whilst for the remaining three the ordering of performance in the jittered and roving conditions was not consistent between subjects within each task, or within subjects across tasks.

In a more recent study, Amitay et al. (2005) also examined the influence of stimulus range on two-interval frequency discrimination. They employed three conditions: fixed (1000 Hz) standard; *roving* standard, where standard frequencies ranged from 900-1100 Hz; and *wide roving* standard, where the standard varied from 570-2150 Hz. Amitay et al. examined the difference limen (or just noticeable difference) averaged across all stimuli in a condition. Difference limens were smallest for the fixed standard and, of interest here, difference limens were smaller (i.e., performance was better) for the wide roving standard than the (narrower) roving standard condition. Amitay et al. interpreted this result in terms of attentional bands by suggesting that in their wide-roving condition participants monitored several frequency bands and used the presentation of the standard tone as a cue to attend to the appropriate listening band, but that in their narrow-roving condition the frequencies may have been too close together to permit this strategy. Instead, the narrow-roving standard led participants to use a single, widened band with a subsequent loss of resolution. Amitay et al. further suggest that in Jesteadt and Bilger's (1974) study the jittered frequencies varied over a sufficiently narrow range (40 Hz) that the attentional band did not need to be broadened by much, whereas in their roving condition, the range necessitated substantial broadening, resulting in poorer resolution.

Amitay et al.'s (2005) finding that discrimination was better when the standard roves over an intermediate range (200 Hz) than a wide range (1580 Hz) suggests that frequency discrimination is not always improved by widening the stimulus range. However, the attentional framework they adopt to explain this result predicts that using a narrow range of standards (i.e., a few tens of Hz) will lead to performance which is better than when a wide range (i.e., several hundred Hz) is used. Thus their attentional model, along with traditional attentional and memory models and the more general *prima facie* argument that frequency perception will be like intensity perception, predicts that frequency discrimination will be better when the stimulus range is narrow. If, on the other hand, the empirical result reported by Amitay et al. is indicative of a more general pattern of worsening discrimination as the stimulus range is reduced, performance will be better with wide-ranging standards, and alternative theoretical accounts will need to be developed. The current experiment directly addresses this issue.

2. Methods

On each trial, the participant heard a standard tone followed by a comparison tone which was either the same as the standard or fractionally lower, and made a same-different judgment. Each participant completed three conditions: a fixed standard condition, a narrow roving condition, in which the standard ranged over 30 Hz, and a wide roving condition, in which the standard ranged over 1500 Hz. In all three conditions, a 1000 Hz tone was used as the middle standard, allowing us to examine the effect of stimulus range on discrimination at a particular frequency. This approach seems preferable to averaging over the different standards used in each condition (Amitay et al., 2005) because it separates the effect of stimulus range from the specific frequencies employed. This might be important if, for example, there were a departure from Weber's law so that sensitivity depended upon stimulus level (e.g., Berliner & Durlach, 1973).

2.1 Participants

Nine participants with experience of auditory psychophysical experiments took part; 8 were paid £30 for participating, the other was author W.M.

2.2 Stimuli

All tones had a total duration of 1000 ms and were gated on and off with 50 ms cosinusoidal ramps at beginning and end. They were generated at a sampling rate of 44.1 kHz and played diotically over Sennheiser eH2270 and HD265 headphones at approximately 80 dB. In the single standard (SS) condition, the standard was a 1000 Hz tone. In the wide-roving condition (WR), three standards were used with frequencies of 500, 1000, and 2000 Hz (i.e. the frequency of each standard was a factor of 2.0 greater than the previous one so that they were evenly spaced on a logarithmic scale). In the narrow-roving (NR) condition, the standard tones had frequencies of 985.2, 1000 and 1015 Hz (i.e. separated by a factor of 1.015). The difference between the standard and comparison tones, Δf, varied from 0.3% to 1.0% for different subjects; the difficulty was selected based upon their previous performance in frequency discrimination tasks and was intended to match approximate overall performance levels across participants.

2.3 Design and Procedure

Each participant completed seven sessions, one in condition SS and three in WR and NR. Each session consisted of 4 blocks of 60 trials. The three NR sessions were grouped together, as were the three WR sessions, giving a total of 6 possible condition orders. Participants completed the sessions over the course of a few days, sometimes completing two or more sessions back to back with a short rest between. Trials from the first block of each session were treated as warm-up and excluded from the analyses. On each trial participants heard the standard tone. After a one second interval they heard the comparison tone and were asked to indicate whether it was the same as the standard or different, and were informed that, if different, the second tone would be slightly lower. They were provided with on-screen feedback about the accuracy of their response for 1s and there was an additional 1s interval before the start of the next trial. The timing of stimulus presentation and response collection was controlled by DMDX (a freely available program for presenting stimuli with millisecond accuracy, Forster & Forster, 2003).

3. Results

Trials on which the participant successfully detected a difference between standard and comparison tones were denoted hits; trials on which the participant correctly identified no difference between standard and comparison tones were denoted correct rejections. The proportion of hits and correct rejections for each level of the standard tone in each condition was used to calculate *d'* as a measure of frequency discrimination. The results are shown in Table 1.

To examine the effects of stimulus range on discrimination at a given frequency (e.g., Harris, 1952), we compared performance in the three conditions when the 1000Hz tone was used as the standard. (An alternative approach is to measure performance in the NR and WR conditions by averaging the d' values for the different standard tones, and to compare these averages [e.g., Jesteadt & Bilger, 1974]. This approach yielded exactly the same pattern of results.) Preliminary analyses established that neither block order nor difficulty (Δf) influenced the differences between the SS, NR and WR conditions. A repeated measures ANOVA revealed a significant effect of condition, $F(2,16) = 17.7$, $\eta_p^2 = .688$, $p < .001$. Paired samples *t*-tests (Bonferroni corrected) indicated a significant difference in the *d'* values between the SS and NR conditions ($t(8)=5.46$, $p < .001$) and between the NR and WR conditions, $t(8) = 4.71$, $p = .004$, but not between the WR and SS conditions [$t(8) = 2.35$]. Inspection of the data from individual participants revealed that, for 8 of the 9 tested, frequency discrimination was worse when the standard roved over a narrow range than when it roved over a wide range. Finally, a repeated measures ANOVA was used to examine the effect of condition on response bias, *c*. The conditions did not differ (F<1).

4. Discussion

The present experiment demonstrates that frequency discrimination with a roving standard is better when the range of stimuli is wide (500-2000 Hz) than narrow (985-1015 Hz). This disagrees with Jesteadt and Bilger (1974) who claimed that performance with a standard which roved over a 40 Hz range was as good as that with a single standard. However, as noted above, Jesteadt and Bilger's data do not convincingly demonstrate a systematic effect of stimulus range on frequency discrimination. Given the highly significant and consistent finding in the current study that frequency discrimination is better when the standard roves over a wide range (1500 Hz) than a narrow range (30 Hz), and the finding by Amitay et al. (2005) of better performance when the standard ranges over a wide range (1580 Hz) than an intermediate range (200 Hz), we suggest that, contrary to what has previously been thought, roving-standard frequency discrimination is not improved by decreasing the stimulus range. This result is unexpected and hard to reconcile with a number of existing theories of frequency discrimination. In what follows we briefly discuss possible explanations for this finding.

4.1 Memory

As noted in the Introduction, it has been suggested that Durlach and Braida's (1969) highly influential model of intensity perception may be extended to describe frequency perception. This theory asserts that, with a fixed temporal interval between tones, performance should be a decreasing function of stimulus range. The results reported here, and those of Amitay et al. (2005), show the opposite pattern, and it is hard to see how the concept of context-coding memory noise could be modified to accommodate our findings.

Durlach and Braida's (1969) theory is not the only memory-based model to have been applied to frequency discrimination. Other workers (e.g., Massaro, 1970) have emphasized the importance of interference and trace decay to discrimination performance, but these similarly fail to account for the current results. Siegel (1972), for example, pointed out that as the number of tones in the stimulus set increases, the length of time and the number of trials intervening since the last presentation of the current tone also increase. Whilst this implies that increasing the number of standard tones will reduce discrimination accuracy, it does not predict any effect of increasing the *range* of the standard tones. Similarly, it is unlikely that the current results can be explained in terms of proactive interference (where the retrieval of more recently presented stimuli is impaired by memories of earlier items); evidence from Ruusuvirta (2000) suggests that in same-different tasks like the one used here proactive interference will only influence response bias, not overall accuracy.

4.2 Attention

When the standard roves over a range of frequencies, participants may employ various attentional strategies but we would typically expect performance in the wide-roving condition to be, at best, the same as in the narrow-roving condition. The current results are therefore problematic for attentional theories. As noted above, Amitay et al. (2005) have tried to explain the finding that performance is better with a wide-roving standard by suggesting that participants use a mixed strategy: in the wide-roving condition, participants monitor several bands and use the standard as a cue to direct attention to the relevant band, whereas in the narrow-roving condition participants broaden a single band to encompass all of the relevant frequencies. In the narrow roving condition of the current experiment, the standards only covered a 30 Hz range. As Amitay et al. note, it seems unlikely that there would need to be much broadening of the listening band to cover this range, or that such broadening would lead to performance that is so much worse than in the wide-roving condition. Furthermore, in the current experiment both standard and comparison tones were played for a full second, long enough that one might expect the participant fully to orient attention to the standard's frequency irrespective of whether he or she is monitoring a single broadened band or several separate channels. Although it is possible that participants use a strategy like that outlined by Amitay et al., the present data necessitate the assumption that even slight widening of the attentional band leads to a dramatic loss of resolution. Thus, while it is undoubtedly the case that attention is an important determinant of frequency discrimination (e.g., Demany, Montandon, & Semal, 2004), it is difficult to develop a convincing attention-based explanation for the current findings.

4.3 Perceptual anchors and plasticity

Conventional memory- and attention-based models struggle to accommodate the effects of stimulus range shown in the current experiment and in the experiment of Amitay et al. (2005). We therefore consider an alternative theoretical orientation which emphasizes learning about the standard tones over successive trials. As some researchers have pointed out (e.g., Ahissar, Lubin, Putter-Katz, & Banai, 2006), when the same standard is used on every trial, participants can form a stable trace of that tone across trials. That is, repeated presentation of the same tone allows formation of a perceptual anchor such that the comparison tones are judged against this anchor rather than against the single presentation of the standard tone on the current trial. In the multiple standards condition the presence of other standards, and the increased number of trials between successive presentations of each, will make it harder to form a perceptual anchor for the three different standards. (Note that we use the term perceptual anchor in the sense of a stable representation of each tone built up over successive trials, as in Ahissar et al. (2006). In a later development of their theory of intensity perception, Durlach and Braida (Braida, Lim, Berliner, Durlach, Rabinowitz, & Purks, 1984) use the same term to refer to the stimuli at the edge of the range; this modified theory can no more readily accommodate the present results than can the original theory.)

As it stands, this account is similar to the memory explanation of poor discrimination performance in one-interval paradigms developed by Siegel (1972) and makes no clear prediction regarding the effect of stimulus range. However, a consideration of the possible neural mechanisms underlying anchor formation does suggest an effect of stimulus range. Each neuron in the primary auditory cortex responds to a range of frequencies, with a peak response to a specific characteristic frequency. There is a growing appreciation that this tuning is somewhat plastic (see Weinberger, 2004, for a review). For example, in conditioning studies training produces systematic changes in the frequency receptive fields (RFs) of neurons in the primary auditory cortex, such that the RF tuning shifts away from the original characteristic frequency towards the frequency of the trained tone (e.g., Bakin & Weinberger, 1990). It seems plausible that a similar process subserves the formation of long-term representations (i.e., perceptual anchors) corresponding to the standard tones used in experiments like the one reported here. That is, we suggest that anchor formation involves retuning/recruiting neurons which normally respond maximally to nearby frequencies to the standard tone's frequency. When the standards are widely spaced, the anchor formation will involve distinct populations of neurons. However, as the standards are moved closer together, there will be an increasing overlap in the neural populations used to represent each one, and a resulting decrease in the fidelity of the anchor. That is, neurons with nearby characteristic frequencies cannot be recruited because they are already being used. As Ahissar et al. (2006) have noted, the failure to form a perceptual anchor may markedly impair discrimination performance. Thus, we suggest that when the standards occupy a narrow range of frequencies it is harder to form a long term trace of each because the representations involve extensively overlapping neural populations, and there is a resulting loss of frequency discrimination.

One advantage of this account is that it may provide an explanation for the difference between the current results and those for intensity discrimination, where increasing the stimulus range reduces accuracy. In intensity discrimination, for quiet and moderate level stimuli, the overall level of activity within a single population of neurons must be compared. Recruitment of neurons with nearby characteristic frequencies will increase the size of the population. Whether the intensity of the standard is kept fixed, varies over a narrow range of intensities, or varies over a wide range of intensities, the increased population size should benefit discrimination performance. So, to a first approximation, recruitment should not differentially affect performance in the different conditions. However, for louder stimuli, when all of the on-frequency neurons are saturated (i.e., firing maximally), loudness information is coded in the spread of activation to off-frequency neurons (Moore & Raab, 1974). If off-frequency neurons are recruited, the population of off-frequency neurons in which the spread of excitation is found will be reduced, and should reduce discrimination performance for louder stimuli. Because the wide-range condition includes more loud stimuli than the other conditions, the wide range condition will be most affected by recruitment.

Throughout this article we have used the term roving standard to indicate a standard selected from a fixed set of unchanging frequencies, spread across a particular range (e.g., Amitay et al., 2005). An alternative way to generate roving standards is by randomly selecting frequencies from a particular range; such a condition might be termed *truly roving*. The use of truly roving standards might provide a way to test the perceptual anchor explanation outlined above. When the standard on each trial is a randomly selected frequency, there will presumably be no opportunity for neural recruitment and the formation of long-term representations of each stimulus. As such, a narrow roving condition will no longer be more difficult than a wide roving condition. Indeed, with a truly roving standard performance may well be better when the stimulus range is narrow, as a wider range will require greater shifts of attention from trial to trial. A future comparison of frequency discrimination in wide-range and narrow-range conditions of an experiment with a truly roving standard may therefore provide a useful way to test the generality of the result reported here, and of the model we propose to explain that result.

5. Conclusions

The key message of the present study is that two-interval frequency discrimination is worse when the standard roves over a narrow range than a wide range. Although we have argued that this pattern is difficult to reconcile with accounts based on memory noise or attention, it may be possible to modify these theories to accommodate the current result. Similarly, our suggestion that the result may be explicable in terms of learning and plasticity may prove incorrect. In either case, the present work, in conjunction with that of Amitay et al. (2005), provides an important empirical finding which demands explanation.

Acknowledgments

This work was supported by ESRC grant RES-000-23-1372. We thank Sygal Amitay and Karolina Kluk for helpful discussion.

References

Ahissar, M., Lubin, Y., Putter-Katz, H., and Banai, K. (**2006**). “Dyslexia and the failure to form a perceptual anchor,” Nat. Neurosci. **9**, 1558-1564.

Amitay, S., Hawkey, D. J. C., and Moore, D. R. (**2005**). “Auditory frequency discrimination learning is affected by stimulus variability,” Perception Psychophys. **67**, 691-698.

Bakin, J. S., and Weinberger, N. M. (**1990**). “Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig,” Brain Res. **536** 271-286.

Berliner, J. E., and Durlach, N. I. (**1973**). “Intensity perception. IV. Resolution in roving- level discrimination,” J. Acoust. Soc. Am. **53**, 1270-1287.

Berliner, J. E., Durlach, N. I., and Braida, L. D. (**1977**). “Intensity perception. VII. Further data on roving-level discrimination and the resolution and bias edge effects,” J. Acoust. Soc. Am. **61** 1577-1585.

Botte, M-C. (**1995**). “Auditory attentional bandwidth: Effect of level and frequency range,” J. Acoust. Soc. Am. **98** 2475-2485.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). “Intensity perception. XIII. Perceptual anchor model of context coding,” J. Acoust. Soc. Am. **76** 722-731.

Demany, L., Montandon, G., and Semal, C. (**2004**). “Pitch perception and retention: Two cumulative benefits of selective attention,” Perception Psychophys. **66**, 609-617.

Durlach, N. I., and Braida, L. D. (**1969**). “Intensity perception. I. Preliminary theory of intensity resolution,” J. Acoust. Soc. Am. **46**, 372-383.

Forster, K. I., and Forster, J. C. (**2003**). “DMDX: A Windows display program with millisecond accuracy,” Behav. Res. Methods Instrum. Comput. **35** 116-124.

Harris, J. D. (**1952**). “The decline of pitch discrimination with time,” J. Exp. Psychol. **43**, 96-99.

Jesteadt, W. and Bilger, R. C. (**1974**). “Intensity and frequency discrimination in one- and two-interval paradigms,” J. Acoust. Soc. Am. **55**, 1266-1276.

Massaro, D. W. (**1970**). “Forgetting: Interference or decay?,” J. Exp. Psychol. **83**, 238-243.

Macmillan, N. A., and Schwartz, M. (**1975**). “A probe-signal investigation of uncertain-frequency detection,” J. Acoust. Soc. Am. **58** 1051-1058.

Moore, B. C. J., and Raab, D. H. (**1974**). “Pure-tone intensity discrimination: some experiments relating to the “near-miss” to Weber’s law,” J. Acoust. Soc. Am. **55** 1049-1054.

Ruusuvirta, T. (**2000**). “Proactive interference of a sequence of tones in a two-tone pitch comparison task,” Psychon. Bull. Rev. **7**, 327-331.

Siegel, W. (**1972**). “Memory effects in the method of absolute judgment,” J. Exp. Psychol. **94**, 121-131.

Weinberger, N. M. (**2004**). “Specific long-term memory traces in primary auditory cortex,” Nat. Rev. Neurosci. **5**, 279-290.

Table 1. Experimental Results.

Participant	Δf (%)	d'_{SS}	d'_{NR}	d'_{WR}
1	1.0	2.13	1.01	1.56
2	0.6	2.94	0.57	1.12
3	0.5	2.67	1.68	2.56
4	0.5	4.83	1.82	3.60
5	0.5	1.89	1.31	2.00
6	0.4	2.80	1.82	1.42
7	0.3	3.57	1.35	2.77
8	0.3	2.46	1.70	3.35
9	0.3	2.94	0.80	1.91
Mean		2.91	1.34	2.25
SD		0.87	0.46	0.87