Level dominance for the detection of changes in level distribution in sound streams

Virginia M. Richards, Yi Shen, and Charles Chubb
Department of Cognitive Sciences, University of California, Irvine, 3151 Social Science Plaza, Irvine, California 92697-5100
v.m.richards@uci.edu, shen.yi@uci.edu, cfchubb@uci.edu

Abstract: Sound streams were generated by randomly choosing the levels of tone pips from two different distributions, A and B. Of the 18 tone pips, the first nine were drawn from distribution A and the second nine from distribution B, or the opposite. The listeners' task was to indicate order, A-B or B-A. In two conditions the A and B distributions differed in mean (condition 1) or variance (condition 2). In contrast to an ideal observer, listeners' strategies were consistent across the two conditions. Analyses suggest that listeners relied primarily on the more intense tone pips in making their decisions.

PACS numbers: 43.66.Lj, 43.66.Mk [QJF]
Date Received: April 30, 2013 **Date Accepted:** June 25, 2013

1. Introduction

Changes in the acoustical environment are associated with changes in the pattern of sound intensity as a function of time and frequency. This may reflect changes within a single source (e.g., formant transitions), changes associated with the removal of a sound source, with the addition of a sound source, etc. For example, adding noise to another noise leads to changes in several features of the sound including increases in the mean and variance of the sound's intensity. It is of interest, then, to evaluate listeners' ability to detect changes in the distributions of levels across time. Here this question is addressed using sequences of tone pips. Of the 18 tone pips, nine had levels drawn from one discrete distribution (distribution A) and the other nine sequential tone pips had levels drawn from a different discrete distribution (distribution B). The listener indicated the order, A followed by B or the opposite. The distributions tested were discrete probability mass functions (*PMFs*) of levels derived using draws from nth-order Legendre polynomials such that the A and B distributions were mirror imagines of one another (Chubb *et al.*, 2007). For example, when $n = 1$ (upper central panel in Fig. 1), for one distribution the probability that a tone pip would be assigned a particular level increased linearly from low to high levels, while for the other distribution the odds were the opposite, i.e., the lowest levels were the most likely to be chosen. Discrete Legendre polynomials were tested because the range of levels available for testing is restricted, or bounded, and Legendre polynomials provide an orthogonal basis of a bounded space.

In addition to estimating sensitivity to differences in the *PMFs*, logistic regressions relating the stimuli tested and listeners' responses estimated relative weights (e.g., Dye *et al.*, 2005) in time and level separately. The resulting relative weights provide an estimate of the relative contribution of tones of different levels and tones presented at different times to listeners' decisions, assuming a linear combination of levels. These relative weights are evaluated relative to two potential models of processing. The first model is an ideal observer model. Because the ideal observer model differs for the different distributions tested (different values of n), one may evaluate whether listener's processing is similar to an ideal observer by tracking the change in relative weights across these two conditions.

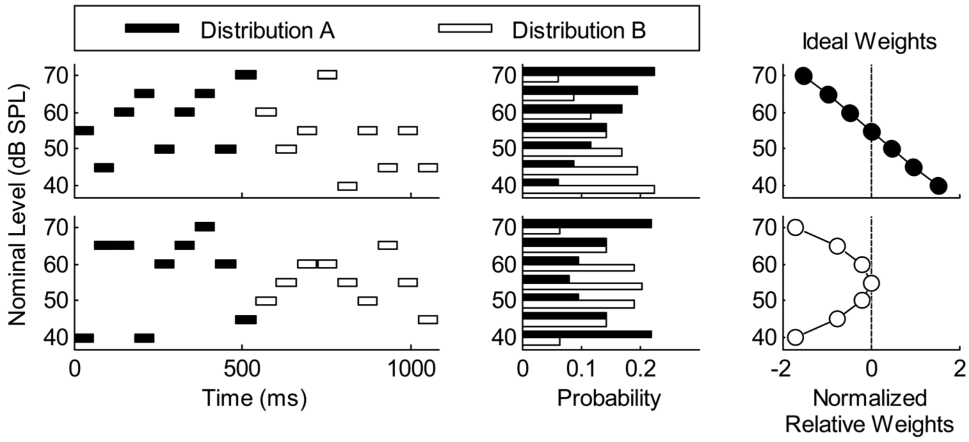

Fig. 1. The left panels show example A-B stimuli plotted as level as a function of time for condition 1 (upper) and condition 2 (lower). The center panels show the *PMFs* for conditions 1 and 2. The right panels shows the weights associated with an observer who differences the level histograms, interval 1 – interval 2. The normalized weights for condition 2 have been shifted to zero, as though the central weight was not estimated (see the Appendix; for condition 1 the expected central weight is zero).

The second model considered is one in which listeners over-emphasize the more intense tone pips in forming their judgments, a model reflecting "level dominance." Unlike the ideal observer model, for the level dominance model listeners rely on the more intense tone pips regardless of the condition being tested. Psychoacoustic studies have demonstrated level dominance across a wide range of studies. An early study by Berg (1990) used a sample discrimination task in which listeners indicated whether the frequencies a sequence of tone pips were drawn from a low-frequency or a high-frequency distribution. Berg found that listeners overemphasized the more intense tones even when they signaled the less reliable samples. Additional results have found level dominance for sequences of tone pips drawn from distributions that differ in mean level (Lutfi and Jesteadt, 2006), when relatively long temporal gaps occur between the tone pips (Turner and Berg, 2007), when identifying object properties that give rise to complex sounds (Lutfi *et al.*, 2008), and for the loudness of sequences of noise bursts (Oberfeld and Plank, 2011). If level dominance is observed when listeners are discerning differences in *PMFs*, it would have significant implications for understanding the perceptual variation evoked by changes in everyday environmental sounds. For example, when a new sound source is introduced or an existing sound source extinguished, listeners may depend only on the more intense sound components reaching the ears to discern those changes, or potentially fail to discern those changes when a source is removed.

To summarize, the goals of the current experiment were to (a) begin to evaluate the mechanisms underlying judgments regarding changes in the statistics within a single sound stream and (b) using relative weights, compare listeners' strategies with two models: an ideal observer and an observer relying predominantly on the more intense sounds.

2. Methods

2.1 Psychophysical methods

Each stimulus was composed of a sequence of 18, 60-ms tone pips with 5-ms cosine-squared onset and offset ramps. The offset of one tone was followed immediately by the onset of the next with no intervening time delay. The sequence of 18 tones was treated as two intervals, each interval being composed of nine sequential tones. The levels of the tones in the first interval were drawn independently and randomly from one distribution, either distribution A or distribution B, and the distributions of tones in the other interval were drawn from the remaining distribution. The listener indicated the order of the two intervals, either A-B or B-A, in a two-interval, two alternative, forced choice procedure. The left panels of Fig. 1 show an example A-B trial for the two conditions tested (top and bottom panels, respectively), plotting level as a function of time.

The central panels of Fig. 1 plot the A and B distributions (black/white bars) that were tested in the first (top) and second (bottom) conditions. These are discrete distributions based on the 1st and 2nd order Legendre polynomials. In condition 1 the *PMFs* were linear, and in condition 2 the *PMFs* were quadratic. Note that in condition 1 the stimuli were drawn from distributions with unequal means but equal variances, and in condition 2 the stimuli were drawn from distributions with equal means but unequal variances. Prior to level rove, the levels of the tone pips ranged from 40–70 dB in 5-dB steps, which will be referred to as "nominal" levels. Two level roves were applied. First, at the outset of each session the calibration level was chosen at random from a 10-dB range. Second, on each trial the overall stimulus level was amplified/attenuated using a range of 15 dB. The roving level reduced listeners' ability to use levels heard on the last trial from influencing their decisions on the current trial. In an effort to reduce the likelihood that forward masking would influence the results, and yet to encourage the perception of a smooth stream, the frequency for each tone pip was chosen at random from a 1/6th octave range. Three frequency regions were tested; [455–510]; [1060–1189]; and [2470–2772] Hz. These will be referred to as the 475-, 1100-, and 2600-Hz frequency regions.

The stimuli were generated digitally using a sampling frequency of 48 000 Hz on a PC, presented diotically to the listeners via a 24-bit sound card (Envy24 PCI audio controller, Via technologies, Inc.), a programmable attenuator and headphone buffer (PA4 and HB6, Tucker-Davis Technologies, Inc.) and a Sennheiser HD410 SL headset. Listeners indicated their order decisions, A-B or B-A, by pressing one of two buttons on a keypad. Each trial was followed by visual feedback indicating the correct response. The experiment was conducted in a double-walled sound-proof booth.

Five listeners participated ranging in age from 20–23 yr, except L4 who was 43. Participants had absolute thresholds of 20 dB hearing level (HL) or better between 250 and 8000 Hz, except the left ear of L4 had an absolute threshold of 25 dB HL at 6000 Hz.

The protocol was as follows. Practice and data collection at one frequency was completed before moving on to the next frequency. The order in which the frequencies were tested for L1 and L2 was {1100, 475, 2600} Hz, and for L3–L5 the orders were {2600, 1100, 475} Hz, {475, 2600, 1100} Hz, and {475, 1100, 2600} Hz, respectively. At the outset the task was explained to the listeners using figures similar to the left and central panels of Fig. 1. Listeners began by practicing condition 2 using the first frequency to be tested. After ten blocks of 50 trials, condition 1 was tested for 10 blocks. This procedure iterated until the listener was above chance levels of performance in both conditions. Before starting a new condition, the listeners were reminded of the new condition's contingencies, either verbally or using figures, depending on the listener's expertise.

After initial practice, data collection began in which 15 blocks of 50 trials were tested in each condition, one after another. The order in which the conditions were tested was randomly drawn for each subject.

When data collection for the initial frequency was finished, the practice/data collection protocol was repeated for the next frequency, etc. A one-tailed t-test was used to test for practice effects, but no practice effects were evident. Following data collection the two subjects (S1 and S3) demonstrating the best sensitivity in condition 2 were tested using discrete distributions based on 3rd-order Legendre polynomial (roughly sideways "s") using tones in the 1000-Hz frequency region. Pilot data suggested differences in distributions based on 4^{th}-order Legendre were not discriminable and so were not tested.

2.2 Analyses

For each listener at each frequency and condition, 700 trials contributed to the summary data, i.e., the first block of 50 trials was discarded. The value of d' for each condition at each frequency region formed the primary data. Additionally, two logistic regressions relating the stimuli and listeners' responses provided estimates of relative weights (e.g., Dye *et al.*, 2005) in level and in time. The resulting regression coefficients were scaled to have a root mean square (RMS) value of 1 for each regression. The sign of the relative weights, negative vs positive, are defined as follows: An A-B response is coded as "0" and a B-A response is coded as "1." Thus, positive coefficients indicate contributions to a B-A response.

To estimate the relative weights associated with level, a two-step procedure was followed. First, for each trial, two histograms indicating the number of tone pips at each of the seven nominal levels were constructed for the first interval and second intervals. Then, the two histograms were differenced bin by bin, interval 1 minus interval 2. The result was a "difference histogram," a vector of the number of tones, interval 1 minus interval 2, at each of the seven levels. The difference histograms summed to zero; therefore the seven level bins provided only six degrees of freedom. To maintain a regression with six degrees of freedom, the central bin of the difference histogram was not used in the analysis. As described in the Appendix, relative weights estimated at the remaining six levels may be interpreted as their contributions to the listener's response *relative to the central bin*.

Next, consider the relative weights in level for the ideal observer. For condition 1 a "B-A" trial is expected to produce a difference histogram with positive values at low levels and negative values at high levels. Therefore, positive coefficients are expected at low-level bins and negative coefficient at high-level bins. Because the magnitude of the expected difference histogram is largest for the lower and higher bins, and zero at the middle bin, the expected relative weights are roughly proportional to the difference histogram (central panel of Fig. 1, B − A). A similar argument leads to the expectation that for condition 2 the expected relative weights would form an upside-down "U," i.e., the difference between histograms B minus A. Patterns of relative weights associated with an ideal observer that compares levels across intervals are plotted in the right panels of Fig. 1 for conditions 1 (upper) and 2 (lower). Reproducing the analysis methods used for the human data, the relative weights have been shifted to have a value of 0 at the central level bin (see the Appendix). Looking forward to the statistical analysis of psychophysical data, one prediction is that relative weights obtained in conditions 1 and 2 would, for an ideal observer, lead to a level-by-condition interaction.

The estimation of the relative weights in time was as follows. For each trial the 18 levels associated with eighteen tone pips formed a vector across both intervals, and a logistic regression relating those values and listeners responses of "0" and "1" indicating A-B and B-A responses, respectively, was run. For each listener, frequency, and condition the coefficients were scaled to have an RMS value of 1.

As with the relative weights in level, the relative weights for an ideal observer in time can be reasoned. For example, in condition 1 if the first tone pip is intense, it would be expected to contribute to a "0" response (A-B trial) whereas a low-intensity tone pip would be expected to contribute to a "1" response (B-A trial). Thus, a negative coefficient between the intensity of the sound in the first time bin and the response is expected, as well as for all of the first nine time bins. For the last nine time bins the opposite is predicted—equal-magnitude positive coefficients. Thus a step function is expected. For condition 2, however, a linear relation between intensity and responses is *not* expected. This reflects that fact that tone pips with relatively high *or* low intensities in the first nine time bins are both predicted to lead to a response of "0." This non-linear decision rule, and its inverse for the last nine time bins, cannot be predicted by a generalized linear model, and so coefficients would not be reliably different from zero. Should this occur, a second step would be to estimate the relative weights using a quadratic model in place of the linear model.

The statistical analyses of the data are in the form of within-subjects analyses of variance (ANOVAs). Even though none of the tests led to the rejection of the hypothesis that sphericity should be rejected (Mauchly's test), the conservative criteria associated with the Greenhouse–Geisser correction was used where available.

3. Results and discussion

The left panel in Fig. 2 plots the averaged values of d' for conditions 1 (filled) and 2 (unfilled) for each frequency averaged across frequencies and the five listeners. Although sensitivity was superior for condition 1 compared to condition 2, the values of d' were relatively low overall. A within-subjects, repeated measures ANOVA indicated a significant effect of condition [$F(1,4) = 17.6$, $p < 0.015$) but not of frequency ($p > 0.15$), nor was the interaction term significant ($p > 0.5$). The two listeners with the highest values of d' in condition 2 (S1 and S3, $d' = 0.70$ and $d' = 0.61$, respectively) were tested using distributions based on 3rd-order Legendre polynomials which yielded even lower values of d' (0.16 and 0.30, respectively).

The central panel in Fig. 2 plots the relative weights averaged across frequencies and listeners. These values are relative to the middle level, which is indicated using an arrow. The patterns of weights can be compared to those expected for a hypothetical listener who ideally differences the level histograms (right panels of Fig. 1, except that level is plotted on the abscissa in Fig. 2). For the human data (Fig. 2), the rate of change of the relative weights as a function of level is more rapid at high than low levels, and there is no clear difference in the relative weights depending on the condition. This is inconsistent with expectations for an ideal observer but consistent with expectations for listeners relying predominantly on the higher-level tone pips, or level dominance.

A within-subjects, repeated measures ANOVA applied to the relative weights indicated a significant effect of level [$F(5,16) = 72.8$, $p < 0.001$) but not of condition ($p > 0.5$) or of frequency ($p > 0.2$). The condition-by-level interaction approached significance [$F(5,20) = 3.4$, $p > 0.075$] but the remaining interactions did not ($p > 0.25$). The near-significant condition-by-level interaction term appeared to reflect the slightly more "bowed" function associated with condition 2 compared to condition 1 (consider lowest level, middle panel). Regardless of whether one prefers to conclude that the observed relative weights for conditions 1 and 2 are shaped somewhat differently, the relative weights in both conditions are reasonably consistent with a model associated with level dominance, and inconsistent with expectations for an ideal observer (right panel of Fig. 1 vs central panel of Fig. 2, noting the rotation of the panels).

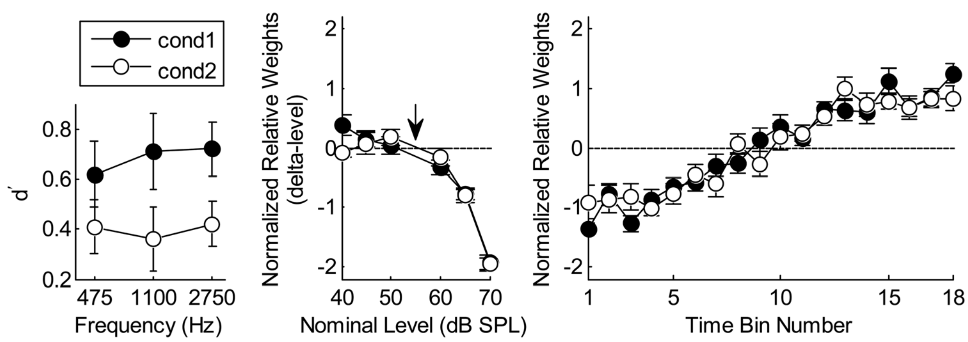

Fig. 2. The left panel shows values of d' averaged across frequency and five listeners for conditions 1 (filled) and 2 (unfilled). Error bars indicate the standard error of the mean. The center and right panels show relative weights averaged across frequency and listeners for level and time, respectively. Conditions 1 and 2 are plotted using filled and unfilled symbols, respectively, and the error bars are standard errors of the mean.

The right panel in Fig. 2 shows relative weights as a function of time (1–18 tone pips as bins) averaged across frequencies and listeners. The pattern indicates a strong dependence on the initial and latter portions of the stimulus and a relatively weak dependence on the more central portions for both conditions. This result might reflect a variety of components of the listeners' decision processes: (a) primacy and recency memory effects on listeners' judgments, or at least the onset and offset of the sound (e.g., Oberfeld and Plank, 2011) (b) a relatively long-duration temporal integrator that effectively smoothes an underlying step function, and/or (c) uncertainty regarding the boundary between the two intervals leading to a de-emphasis of the stimuli presented near the middle of the sound stream. The pattern of relative weights was similar across conditions, in contrast to the predictions of the ideal observer analyses of a step function in condition 1 and unreliably estimated relative weights from zero in condition 2.

A within-subjects, repeated measures ANOVA applied to the time-relative weights indicated a significant effect of time [$F(17,68) = 30.3$, $p < 0.001$] but not of condition ($p > 0.75$), of frequency ($p > 0.40$), or any of the potential interactions ($p > 0.2$). These time-relative weights, like the level-relative weights, suggest that listeners rely on a single mechanism/decision strategy to determine the order of A and B regardless of condition.

These results suggest that when asked to indicate the order of statistical changes—from a low mean to a high mean, and from low variance to high variance, or the opposite order—listeners rely on a single mechanism/strategy, and that the strategy is the same across frequency regions. The underlying mechanism/strategy is not consistent with the predictions of an ideal observer but is largely consistent with level dominance. The results also indicate that listeners did not equally weight information from all of the first nine, or all of the second nine, tone pips equally in making their decisions. Instead, tones close to the center portion of the sequence are under-weighted relative to the early and late tone pips. That is, there is no evidence listeners form a contrast between tones just before and just after the interval boundary, which might be predicted if a salient boundary between the two sets of tone pips had been perceived. Potentially, separating the two intervals with a period of silence (as is typical in two-interval, two alternative forced choice procedures) would have led to a pattern more similar to that expected for a contrast in that the ending of the first interval would be associated with a larger negative weight and the onset of the second interval would be associated with a larger positive weight.

Additional computer simulations indicate that, had listeners maintained the same relative weights in level observed for conditions 1 and 2, sensitivity to changes in the order of A and B would be near chance levels for *PMFs* derived from Legendre polynomials with an n of 3, a result observed for our two most sensitive listeners. This "null result," in addition to the positive results described above, suggest that level dominance may play an important role in the perceptual organization of the acoustical information reaching the listeners' ears. Acknowledging that the current experiment evaluated circumscribed changes in a sparse acoustic stream, the results indicate that at least in the absence of pitch cues, the perception of changes in the acoustic stream associated with variations in the acoustics of a single source and changes in the number of sound sources may be dominated by the more intense elements of the sound reaching listeners' ears.

Acknowledgments

This work was supported by Grant No. R21 DC010058 from NIDCD. We acknowledge Theodore Lin and Andrew Silva who assisted with data collection and analysis.

Appendix

Here expected values for the relative weights for the "difference histograms" are considered. For ease of explanation, assume there are only three potential levels, L_1, L_2, and L_3, whose expected values are symmetric about L_2 as holds for distributions A and B in the current experiment. First a histogram for the series of tones from interval 1 is formed, and then the histogram for the series of tones from interval 2 is formed. Each histogram has three elements indicating the number of tone pips with the three potential levels. Let X_i, $i = 1$–3, be the difference in the number of tone pips at each level, interval 1 minus interval 2, for the levels L_1, L_2, and L_3. By construction $\Sigma_i X_i = 0$.

Consider only the linear portion of the logistic regression (e.g., ignoring the transformation to 0–1): $y = aX_1 + bX_2 + cX_3 + d$. Because $\Sigma_i X_i = 0$, $X_2 = -X_1 - X_3$, $y = aX_1 - bX_1 - bX_3 + cX_3 + d$ or $y = (a - b)X_1 + (c - b)X_3 + d$. Accordingly, if the coefficients for all but one of the coefficients, b in this example, is estimated the resulting coefficients, $(a - b)$ and $(c - b)$, are shifted by the value of the coefficient associated with the missing random variable.

References and links

Berg, B. G. (**1990**). "Observer efficiency and weights in a multiple observation task," J. Acoust. Soc. Am. **88**, 149–158.

Chubb, C., Nam, J.-H., Bindman, D. R., and Sperling, G. (**2007**). "The three dimensions of human visual sensitivity to first-order contrast statistics," Vision Res. **47**, 2237–2248.

Dye, R. H., Stellmack, M. A., and Jurcin, N. F. (**2005**). "Observer weighting strategies in interaural time-difference discrimination and monaural level discrimination for a multi-tone complex," J. Acoust. Soc. Am. **117**, 3079–3090.

Lutfi, R. A., and Jesteadt, W. (**2006**). "Molecular analysis of the effect of relative tone level on multitone pattern discrimination," J. Acoust. Soc. Am. **120**, 3853–3860.

Lutfi, R. A., Liu, C. J., and Stoelinga, C. N. J. (**2008**). "Level dominance in sound source identification," J. Acoust. Soc. Am. **124**, 3784–3792.

Oberfeld, D., and Plank, T. (**2011**). "The temporal weighting of loudness: Effects of the level profile," Atten. Percept. Psychophys. **73**, 189–208.

Turner, M. D., and Berg, B. G. (**2007**). "Temporal limits of level dominance in a sample discrimination task," J. Acoust. Soc. Am. **121**, 1848–1851.