Laboratory 3 Report:
Pitch shift of pure tones induced by masking noise

Alex Latham [alatham@mit.edu]

May 11, 2016

Abstract

This experiment measures the influence of masking noise on pitch discrimination. It has been noted (Fastl 1971, Stoll 1985) that pure tones can be perceived as being of different pitches when presented with masking noise and without masking noise. This lab will use an adaptive procedure to find a listener's perceived phase shift of a tone when presented with various levels of Gaussian noise. The results of this lab show that pitch shift occurs significantly at noise levels from 30 to 60 dB, and the relationship between pitch shift and noise level over this range can be approximated as linear.

1 Introduction

1.1 Background

The pitch of an observed tone is primarily determined by the tone's frequency. However, other factors can influence a listener's perception of pitch, including the presence of various maskers. For example, when a listener hears a tone presented simultaneously with a masking noise, the pitch of the tone is perceived to be different than the same tone presented alone. The amount by which the pitch of a tone changes in the presence of a masker is called *pitch shift*, and is given by

$$PS = (f_a - f_T)/f_T \qquad \text{(Eq. 1)}$$

measured as a percentage of the base tone f_T (the tone presented with a masker), where f_a is the frequency of the unaccompanied tone the listener perceives to correspond to the tone in noise. Varied noise levels can cause different amounts of pitch shift for a given tone, and pitch shift also varies considerably interpersonally (Stoll, 188). The relationship between pitch shift and noise level can help us understand the results of other psychoacoustic experiments and has implications in music and other applications, for example in determining tolerable thresholds for noise levels before pitch perception becomes distorted.

This lab will extend experiments testing the effects of masking noise on pitch shift conducted by Stoll (1985). Stoll tested the effect of noise on both pure tones and harmonic residue tones, which are implicit tones suggested by a few of the tone's harmonics rather than the tone itself. Stoll tested using low-pass noise cut off just above the frequency of the tones tested, and allowed listeners to adjust a tone to 'match' a test tone presented in noise. His experiment determined that amid noise, tones will be perceived as being of higher pitch than without, with pitch shifts noted up to 6%. Additionally, as the noise level of the masking noise is increased, pitch shift becomes more dramatic. Results from Stoll's tests on pure tones are shown in figure 1.

1.2 Objective

This lab will be testing in particular the effects of various levels of wideband noise on the perceived pitch of a pure tone. Using an adaptive procedure as specified by H. Levitt (1970) rather than the listener-driven procedure used by Stoll, an 1000 Hz tone will be presented amid Gaussian white noise, along with an adapted, unmasked tone, to determine the listener's threshold for pitch discrimination. Gaussian white noise

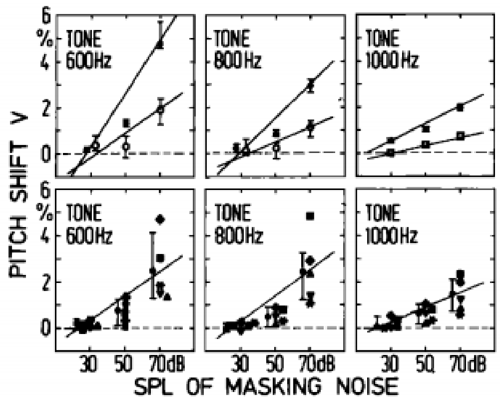


Figure 1: Results from Stoll (1985) for pitch shift of pure tones of frequency 600, 800 and 100 Hz, presented with 30, 50, and 70 dB of masking noise. Data from two subjects is shown on the top row; data from five subjects is shown on the bottom row.

is commonly used to represent and model the background noise and self-interference of various ratio systems, so this choice represents a minor extension of Stoll's tests. An assumption implicit in Stoll's discussion is the linearity of the relationship between noise level and pitch shift- the relative slopes of noise level/pitch shift curves are often compared (Stoll, 189).

This lab will test for (H1) confirmation of the existence and direction of pitch shift for a pure tone in noise, and (H2) the relationship between noise level and pitch shift- if pitch shift in fact increases with noise level, and whether linearity is a reasonable assumption.

2 Measurements

To test the hypotheses motivated by Stoll's experiment, we use an adaptive procedure with experimentally chosen parameters and a variety of features to limit bias. Below is a description of the procedure used in this experiment, and a preliminary description of the results obtained.

2.1 Procedure

The procedure used in this lab is an adaptation of Levitt's adaptive, 2-alternative forced-choice procedure commonly used in psychoacoustic testing. The efficacy of this procedure for use in sound discrimination experiments was explored in Lab 1. In each of two testing sessions, a 50 dB, 1000 Hz tone was presented simultaneously with Gaussian white noise of 30, 50, and 60 dB (all within in the range tested by Stoll). Each trial, a listener was presented with both the fixed, 1000 Hz tone in noise and an adapted tone whose frequency was adjusted based on the listener's responses. The stimuli are separated by a 30 ms silence. The step size of the tone's adjustments was reduced by a factor of the number of presentations upon a correct response, a common strategy for step-size reduction in adaptive procedures. 3 Hz were added to the step size upon an incorrect response, as a way to 'back out' of incorrect convergences. 3 Hz was determined to be just above the discrimination threshold of the listener. The implementation of these rules can be seen more clearly in the Code section.

Testing Parameters	
Tone freq.	1000 Hz
Tone level	50 dB
Noise band	0-22,000 Hz
Noise level	30, 50, 60 dB
# Turnarounds	9
Initial freq.	1040, 960 Hz
Initial step size	20 Hz
Minimum step size	1 Hz

Table 1: Parameters of adaptive testing procure.

The initial frequency of each trial is alternated between 1040 Hz and 960 Hz. This alternation provides protection against upward or downward bias the listener might present. The presence of this bias is explored in section 3.3. The initial step size was chosen to be 20 Hz, which, with a relatively high number of turnarounds (9, before a trial is concluded), produced consistent results in pre-experimental testing. Finally, a minimum step size of 1 Hz was chosen to avoid step sizes decrementing too far below the listener's discrimination threshold. 1 Hz is also under this threshold, and so likely did not have an adverse affect on measurements.

In addition to alternating between initial tones, other measures were taken to reduce listener bias. Each session involved 4 trials at each noise level, with two at each initial frequency, and alternating the order of presentation (the adapted tone, and fixed tone with noise) between each trial. Between sessions, the order of noise levels tested was reversed. Finally, before each session, the listener practiced for 3 trials with a noise level of 50 dB. The listener was not presented with feedback. Table 2 depicts the organization of testing in sessions 1 and 2.

		Session 1			Session 2		
		Noise Level			Noise Level		
Trial	Initial Freq.	30 dB	50 dB	60 dB	60 dB	50 dB	30 dB
1	1040 Hz	[S, N]	[S, N]	[S, N]	[S, N]	[S, N]	[S, N]
2		[N, S]	[N, S]	[N, S]	[N, S]	[N, S]	[N, S]
3	960 Hz	[S, N]	[S, N]	[S, N]	[S, N]	[S, N]	[S, N]
4		[N, S]	[N, S]	[N, S]	[N, S]	[N, S]	[N, S]

Table 2: Organization of each testing session. Each session tested at 30, 50, and 60 db noise (reversed between sessions), and each trial alternated the presentation of the tone and tone with noise. [S, N] represents a trial where the tone was presented before the tone with noise; [N, S] represents a trial where the tone with noise was presented before the unmasked tone. Additionally, the first session's trials were upward biased, and the second session's trials were downward biased.

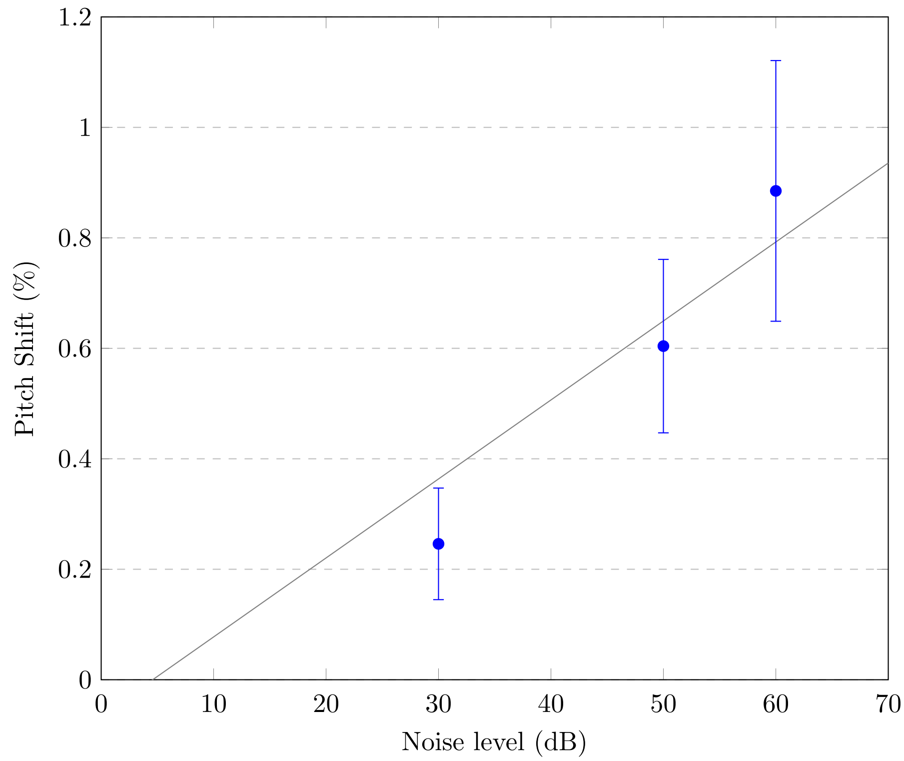
2.2 Results

Table 3 presents the average pitch shifts measured at 30, 50 and 60 dB noise levels. These values are shown graphically with standard errors and best-fit line in figure 1. We immediately note an increase in pitch shift with masking noise level, as well as an increase in standard error with masking noise. The following section contains a complete analysis of these results. Unprocessed data can be found in the Appendix.

Noise Level					
30 dB		50 dB		60 dB	
Ave. PS	SD PS	Ave. PS	SD	Ave. PS	SD
.246%	.101%	.604%	.157%	.885%	.236%

Table 3: Adaptively determined pitch shift of 1000 Hz tone at various noise levels.

Figure 1: Adaptively determined pitch shift of 1000 Hz tone at various noise levels, with standard errors shown. Best-fit line including origin is $pitch\_shift = .0143(noise\_level) - .0655$.



3 Analysis of Results

We will use the measured results to test for the following: (H1) the existence and direction of pitch shift caused by noise, (H2) the linearity of pitch shift with respect to noise level (as assumed by Stoll), and, as an issue of measurement, the listener bias between upward and downward biased tests.

3.1 Existence of Pitch Shift

We can see from figure 1 that, as measured, pitch shift does seem to occur with the introduction of noise. However, to determine whether or not our measurements represent significant pitch shift, we will make the following comparisons- first, between the pitch shift measured at 30 dB and the pitch shift measured at 50 db, and between the pitch shift at 50 dB and the pitch shift at 60 dB. If these measurements are significantly different from each other, and increase or decrease in a consistent direction, we can conclude that pitch shift did consistently occur under our testing conditions.

First, we will determine if the 30 dB measurement represents a significant departure from a hypothetical measurement in the no-noise case. To test this, we will find the theoretical probability of the pitch shift measurement at 30 dB under the null hypothesis- that pitch shift *does not* occur. The average pitch shift measured at 30 dB is .246%, with standard deviation of .101%. So, assuming we expect 0% pitch shift (using the same standard deviation as the 30 dB test), the probability of the 30 dB measurements being .246% (or a less likely value) is 1.46%. So, our measurements for pitch shift at 30 dB are indeed significantly different than theoretical results from a no-noise test.

We now use the Mann-Whitney test to determine whether the pitch shift measurements at each other noise level are significantly different, and in the same direction (consistently greater, or less-than). The Mann-Whitney test is ideal for the data collected- there are relatively few data points, and we have not assumed that our results should follow a normal distribution. First, we will compare the measured pitch shift at 50 dB to that at 30 dB. The results of this test (shown in figure 2) suggest that the pitch shift measured with 50 dB noise is significantly higher than that measured at 30 dB. Likewise, a Mann-Whitney test comparing the pitch shift at 50 dB with that at 60 (figure 3) suggests that the pitch shift at 60 dB is significantly higher than that at 50 dB.

These tests have confirmed that pitch shift does occur significantly in the presence of wideband masking noise and correlates positively with noise level, confirming H1.

3.2 Linearity of Pitch Shift in Noise Level

Throughout his paper, Stoll compares the slopes of various noise level/pitch shift curves to describe differences between various subjects and tone types. Verifying that linearity is a reasonable assumption will validate, or invalidate, Stroll's comparisons, and help us better understand the nature of the relationship between pitch shift and noise level in the range tested.

To determine whether or not the results approximate a linear relationship between pitch shift and noise level, we will validate the fitment of the best-fit line as seen in figure 1. To do this, we examine the adjusted r^2 value of the data to test the 'goodness of fit' of the best-fit line. We will include the no-noise case as a hypothetical data point. Using Pearson's formula

$$r = \frac{\Sigma(xy)}{\Sigma(x^2)\Sigma(y^2)} \qquad \text{(Eq. 2)}$$

	Pitch Shift	
Trial	30 dB	50 dB
1	.145%	.599%
2	.228%	.294%
3	.358%	.476%
4	.141%	.674%
5	.182%	.583%
6	.181%	.757%
7	.287%	.687%
8	.447%	.764%

(a)

$\mathcal{O}$	30 dB or 50 dB	PS
1	30	.141%
2	30	.145%
3	30	.181%
4	30	.182%
5	30	.228%
6	30	.287%
7	50	.294%
8	30	.358%
9	30	.447%
10	50	.476%
11	50	.583%
12	50	.599%
13	50	.674%
14	50	.687%
15	50	.757%
16	50	.764%
	U = 2	

(b)

Figure 2: (a) Average measured pitch shifts of an 1000 Hz tone with noise levels 30 and 50 dB. **(b)** Sorted pitch shifts from 30 and 50 dB noise levels. The Mann-Whitney score of these two sets of results is 2. The chance of this, or a less likely outcome under the null hypothesis is $p = 0.00097$. So, the pitch shift measured at 50 dB noise is significantly higher than that measured at 30 dB noise.

	Pitch Shift	
Run	50 dB	60 dB
1	.599%	1.098%
2	.294%	1.262%
3	.476%	.604%
4	.674%	1.065%
5	.583%	.831%
6	.757%	.863%
7	.687%	.673%
8	.764%	.680%

(a)

$\mathcal{O}$	50 dB or 60 dB	PS
1	50	.294%
2	50	.476%
3	50	.583%
4	50	.599%
5	60	.604%
6	60	.673%
7	50	.674%
8	60	.680%
9	50	.687%
10	50	.757%
11	50	.764%
12	60	.831%
13	60	.863%
14	60	1.065%
15	60	1.098%
16	60	1.262%
	U = 11	

(b)

Figure 3: (a) Average measured pitch shifts of an 1000 Hz tone with noise levels 50 and 60 dB. **(b)** Sorted pitch shifts from 50 and 60 dB noise levels. The Mann-Whitney score of these two sets of results is 11. The chance of this, or a less likely outcome under the null hypothesis is $p = 0.01578$. So, the pitch shift measured at 60 dB noise is significantly higher than that measured at 50 dB noise.

we get $r = 0.968$ and $r^2 = 0.937$. With a sample size of 4 and one predictor variable (noise level), this yields an adjusted r^2 value of 0.91. An adjusted r^2 value of 1 implies a perfect fitment of data, so an adjusted r^2 value of 0.91 suggests that a linear model is largely adequate to describe the relationship between phase shift and noise level for samples in the 30 to 60 dB noise level range.

So, Stoll's implicit assumption of the linearity of this relationship is well founded. Including results from the previous section, we conclude that there is a positive correlation between pitch shift and masking noise level- that is, with an increase in masking noise level, we can expect a roughly linear increase in pitch shift over noise levels 0 to 60 dB. This confirms H2.

3.3 Upward/Downward bias

A final test, though not directly related to our initial hypotheses, is to measure the degree of upward and downward bias depicted by the results. If the listener performed differently in the upward biased tests than the downward biased tests, we might infer either a failure of the procedure or extract information about the nature of masking and pitch shift. Averages for trials separated by bias are shown in table 4.

In the first two trials of each session, the initial tone frequency was 1040 Hz, while in the last two trials it was 960 Hz. We will compare the average pitch shifts of upward biased tests and downward biased tests at each noise level using a Mann-Whitney test. We see that comparing trials 1, 2, 5, and 6 (upward biased) with 3, 4, 7, and 8 (downward biased) for the 30 dB noise level tests produces a Mann-Whitney score of $U = 13.7$, which makes the chance of the observed, or a less likely set of measurements, 52%. For 50 dB tests this probability is 63%, and for the 60 dB tests, 34%. This confirms that the distributions are very similar, meaning that there was no significant upward or downward bias at any noise level.

Table 4: Average threshold frequency and SD for upward and downward biased tests at noise levels 30, 50, and 60 dB. Upward biased trials are 1, 2, 5, and 6; downward biased trials are 3, 4, 7, and 8 (refer to table 5).

	30 dB Noise Level		50 dB Noise Level		60 dB Noise Level	
	Ave. Freq. (Hz)	SD (Hz)	Ave. Freq. (Hz)	SD (Hz)	Ave. Freq. (Hz)	SD (Hz)
Upward biased	1001.84	0.34	1005.58	1.93	1009.28	2.98
Downward biased	1003.08	1.29	1006.50	1.23	1007.56	2.09
Overall	1002.46	1.01	1006.04	1.57	1008.85	2.36

4 Discussion of Results

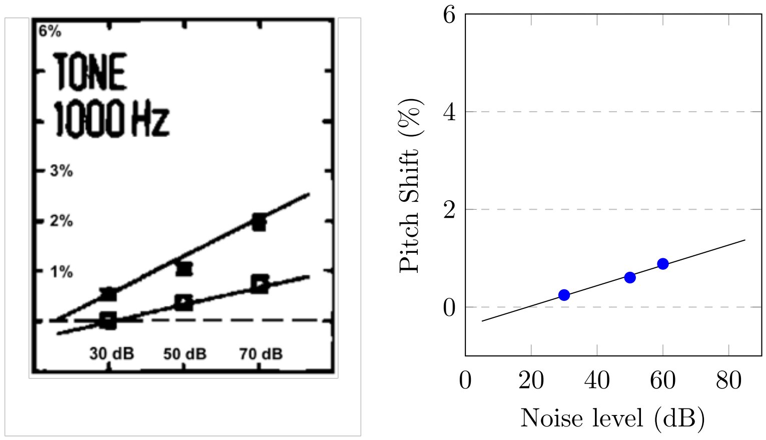
The results obtained in this laboratory have confirmed the hypotheses prompted by Stoll's experiment. In interpreting these results, it is important to understand potential limitations of the procedure and outcomes, as well as to recognize their implications, clinical or otherwise.

4.1 General Notes

One aspect of the results not yet discussed is the increase in standard deviation with noise level as seen in the error bars pictured in figure 1. The standard deviation for trials at 30 dB is .101%, .157% for trials at 50 dB, and .236% for trials at 60 dB. This represents a considerable increase, likely attributable to the nature of pitch shift. At low noise levels, the tests more closely approximate a simple frequency discrimination test. However, as the effect of the masking noise increases, our perception is distorted more significantly. It might be that at increased noise levels, we rely less on absolute matching and more on particular audible cues, which may vary in reliability, particularly when tones are overpowered by their masking noise (such as in the 60 dB case, which has the highest standard deviation). It might also be that our ability to overcome the effects of pitch shift from masking are variable, dependent on listener focus or other conditions. The increase in standard deviation might be noted as a procedural issue, but it is more likely a psychophysical effect- this increase mirrors that seen in Laboratory 2, where the standard deviation of the just-noticeable intensity differences increased significantly with sensitivity.

This experiment faces two fundamental limitations. First is the range of noise levels tested. Though the noise levels in this experiment fall within the range of those tested by Stoll, and likely cover a large span of common noise levels, more noise levels should have been tested (and with finer granularity) to more deeply understand the relationship between pitch shift and noise level. Similarly, the limited number of trials pose an issue. In section 3.3, there was one fewer value per test than recommended to produce accurate Mann-Whitney results. More trials would be necessary to confidently verify the linearity of the relationship between pitch shift and noise level.

Additionally, the procedure used in this experiment differed slightly from Stoll's. First, it was adaptive rather than listener-driven. Stoll allowed listeners to manually adjust the sample tone to the tone in noise. More control might have allowed listeners to more thoroughly compare tones and produce more accurate thresholds. Second, this procedure used Gaussian white noise rather than low-pass noise. Though this may have impacted results, it is more likely that the use of wideband noise represents a successful generalization



(a) **(b)**

Figure 4: A direct comparison of results from this experiment (b) and Stoll's test of an 1000 Hz tone under various levels of masking noise for two subjects (a). Best fit line in (b) has been recalculated to ignore the origin, as Stoll appears to have done. Note, for Stoll's lower subject particularly, the similarity in slope and position of the best-fit line.

of Stoll's experiment. Because Stoll's noise also includes the frequencies of the tones he tested, it is unlikely that the use of Gaussian white noise (which only introduces even lower frequencies) was problematic.

Finally, the parameters for this procedure were determined empirically. The initial step size, initial tone frequency, and step-size-changing rules were chosen to minimize observed problems with becoming 'stuck' at incorrect values and ensuring a sufficient number of trials before convergence. A relatively high number of turnarounds was used (9) to help ensure the accuracy of the final convergence.

In general, the results of this experiment closely follow Stoll's, and our successful validation of the hypotheses represents a strengthening of Stoll's work. A direct comparison of the results obtained in this lab is shown alongside Stoll's results from the same test in figure 4. Though high levels of interpersonal variation mean that a direct comparison is unfounded, there appears to be strong consistency between the results obtained here and by Stoll.

4.2 Implications

We have shown that (H1) pitch shift does occur in an upward direction for tones presented simultaneously with noise and (H2) that a linear relationship successfully approximates that between pitch shift and noise level. These conclusions are in agreement with Stoll, and others who have tested the phenomenon of pitch shift, including Terhardt (1971).

These results suggest that in acoustic systems with background noise or feedback, engineers and audiologists should be cognizant that as noise levels increase, so will the shift in listeners' perception of tones. If undistorted perception of pitch is relevant to a musical or clinical application, a reliable model for acceptable thresholds for pitch shift based on noise level is important, and we have shown that such models can be approximated using this lab's testing procedure.

References

[1] Levitt, H. "Transformed Up-Down Methods in Psychoacoustics." *The Journal of the Acoustical Society of America* (1970): 467-77.

[2] Stoll, G. "Pitch shift of pure and complex tones induced by masking noise" *J. Acoust. Soc. Am.* 77.1 (1985): 188-192.

[3] Terhardt, E.; Fastl, H. "Influence of Masking Tones and Noises on the Pitch of a Pure Tone" *Acta Acustica united with Acustica* 25.1 (1971): 53-61.

5 Appendix

Table 5: Results of adaptive procedure to determine pitch discrimination threshold between 1000 Hz tone and 100 Hz tone in various levels of masking noise. Threshold frequencies are determined from average of last 9 turnarounds.

	Noise Level					
	30 dB		50 dB		60 dB	
Run	Thresh. Freq. (Hz)	# Pres.	Thresh. Freq. (Hz)	# Pres.	Thresh. Freq. (Hz)	# Pres.
1	1001.45	28	1005.99	15	1010.98	18
2	1002.28	24	1002.94	15	1012.62	14
3	1003.58	18	1004.76	18	1006.04	22
4	1001.41	20	1006.74	17	1010.65	15
5	1001.82	21	1005.83	16	1008.31	17
6	1001.81	28	1007.57	15	1008.63	23
7	1002.87	24	1006.87	19	1006.73	20
8	1004.47	19	1007.64	20	1006.80	17
Ave. Freq.	1002.46 Hz		1006.04 Hz		1008.85 Hz	
SD Freq.	1.01 Hz		1.57 Hz		2.36 Hz	

6 Code

The code used to perform the adaptive tests for this lab is presented below. The tests were written in Python and run over a command-line interface. PyAudio is the sound generation package used. Note that print statements offering listener feedback are shown, but were removed for testing.

```
import math
import numpy
import pyaudio
import audioop
import random
import struct
import pylab as pl

FREQ = 1000
DBCONST_TONE = 26.5 #26.5 for 50 dB
DBCONST_NOISE = 2.6 #2.6 for 30db noise; 26.5 for 50 dB; 85 for 60 dB
NUM_TURNAROUNDS = 9
START_FREQS = [1040,960]
ORDER_SN = 1 #1 for (S,N), 0 for (N,S)


def sine(frequency, length, rate):
     length = int(length * rate)
     factor = float(frequency) * (math.pi * 2) / rate
     sine = numpy.sin(numpy.arange(length) * factor)
     return sine

def noise(frequency, length, rate):
    length = int(length * rate)
    factor = float(frequency) * (math.pi * 2) / rate
    noise = numpy.random.rand(length)
    return noise


def play_tone(stream, frequency=440, length=1, rate=44100):
    #length in half-seconds
    chunks = []
    tone = DBCONST_TONE * sine(frequency, length, rate)
    silence = numpy.zeros(30000)
    noiseSound = DBCONST_TONE * sine(FREQ, length, rate) + \
        DBCONST_NOISE*noise(FREQ, length, rate)

    if ORDER_SN:
        chunks.append(tone)
        chunks.append(silence)
        chunks.append(noiseSound)
    else:
```

```
        chunks.append(noiseSound)
        chunks.append(silence)
        chunks.append(tone)
    chunk = numpy.concatenate(chunks) * 0.25
    stream.write(chunk.astype(numpy.float32).tostring())
    return chunk

def get_rms(block):
    s = 0
    for i in block:
        s += i**2
    return math.sqrt(float(1./len(block)) * s)

def test(current):
    p = pyaudio.PyAudio()
    stream = p.open(format=pyaudio.paFloat32,
                    channels=2, rate=44100, output=1)

    play_tone(stream, current)

    stream.close()
    p.terminate()

    if (current >= FREQ):
        if ORDER_SN:
            return 1
        else:
            return 2
    else:
        if ORDER_SN:
            return 2
        else:
            return 1

def dbPlot():
    p = pyaudio.PyAudio()
    stream = p.open(format=pyaudio.paFloat32,
                    channels=2, rate=44100, output=1)

    chunk = play_tone(stream, FREQ)

    t = numpy.arange(len(chunk))*1.0/44100
    a = 20*numpy.log10(numpy.abs(numpy.fft.rfft(chunk)))
    f = numpy.linspace(0, 44100/2.0, len(a))
    pl.plot(f, a)
    print numpy.sqrt(numpy.mean(numpy.square(a)))
    pl.xlabel("Frequency(Hz)")
    pl.ylabel("Power(dB)")
    pl.show()

    stream.close()
```

```
    p.terminate()


if __name__ == '__main__':

    #dbPlot()

    for j in START_FREQS:

        for i in [1,0]:

            ORDER_SN = i

            current = j #1040/960
            step = 20.0 + random.uniform(-5, 5) #initial
            min_step = 1.0
            turnarounds = []
            past_answer = True
            num_steps = 0

            while (len(turnarounds) < NUM_TURNAROUNDS):
                num_steps += 1
                results = test(current)
                if step < min_step:
                    step = min_step
                answer = input('1 if first tone was higher, else 2: ')
                if (int(answer) != results):
                    print "wrong"
                    if (past_answer == True):
                        turnarounds.append(current)
                        past_answer = False
                    if (current<FREQ):
                        current-=step
                    else:
                        current+= step
                    step += 3.0
                else:
                    print "correct"
                    if past_answer == False:
                        turnarounds.append(current)
                        past_answer = True
                    if (current<FREQ):
                        current+=step
                    else:
                        current-= step
                    step /= num_steps

            print "turnarounds: ", str(turnarounds)
            print "Ave: ", str(reduce(lambda x, y: x + y, turnarounds)
                len(turnarounds))
            print "# Pres: ", str(num_steps)
```