Proceedings of the Institute of Acoustics

ON THE AUDIBILITY OF TRANSIENT PITCH SHIFTS

J A S Angus & C M Barnes

Department of Electronics, University of York, Heslington, York

1. INTRODUCTION

The audibility of dynamic pitch shifts affects several areas of sound reproduction from the transient behaviour of playback systems to the problems of acquiring pitch information for synthesiser applications.

Many researchers have investigated the Frequency Difference Limen (FDL) for stationary tones and the Frequency Modulation Detection Threshold (**FMDT**) for dynamic tones. This paper briefly describes the work of previous authors and details a unified approach to analysis of their results. The techniques used to normalise FDL and FMDT data to take into account variations in experimental technique, stimuli and parameters are explained. These include the development of thresholds in terms of *Pitch*, regarded as log frequency. A psychoacoustical experiment intended to test the hypotheses drawn from analysis of the synthesised data is also presented along with its results which are discussed.

2. PITCH DIFFERENCE LIMENS

Many authors have published data on the threshold for detection of a frequency difference between two stationary tones — Frequency Difference Limen (FDL), $\Delta f = f_2 - f_1$. How the FDL varies with the duration of the tones presented is of particular interest as it relates to a sampling theory of frequency modulation detection.

2.1 Previous work

Turnbull[11] observed that the accuracy of pitch discrimination decreased as the tone duration was reduced. More recently Moore[6], investigated FDLs as a function of pure tone frequency for a wide range of tone durations (D). A discontinuity in the FDL behaviour was observed at about 4KHz. In this report only frequencies below 2KHz are considered, for which the FDL increased as the tone duration decreased, for short-duration stimuli.

Mark[5] performed similar experiments with signals consisting of single-, double-, & triple-cycle sinusoids. The FDL of these signals for frequencies in the range 256–2048Hz was found to be in the order of 1–2 semitones. The FDL increased for fewer cycles of the sinusoid.

2.2 Analysis

2.2.1 Normalisation. Results for stimuli at different frequencies can be compared if the data is 'normalised'. This was accomplished by evaluating thresholds as a fraction of the stimuli frequency (relative FDL — $\Delta f/f$), section 2.2.3 shows why this should be a reasonable way of looking at the data. Instead of plotting the thresholds as a function of duration (time) the abscissa was also normalised by using the number of cycles ($\mathcal{N} = f/D$) of the stimuli.

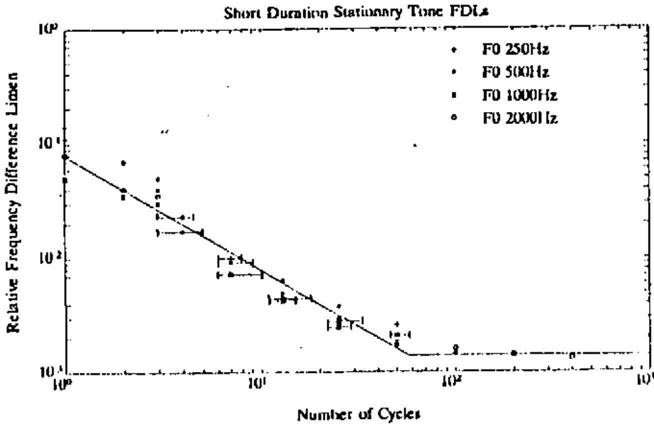

Figure 1: Plot of relative FDL for stationary, pure tones as a function of stimuli number of cycles ($\mathcal{N}$) (log scale on both axes). The exact number of cycles for short-duration tones depends on the allowance made for the rise/fall times, indicated by the horizontal bars. Data for $\mathcal{N} = 1, 2, 3$ Mark[5]; $\mathcal{N} > 3$ Moore[6] subject T.C.. Data trends fit two straight lines as shown (see text).

2.2.2 Interpretation. Normalised FDL results of Moore[6] below 2KHz and of Mark[5] are presented in Fig.1. The data can be fitted by two straight lines. Using three lines[1] would improve the fit in the middle region, but is unnecessary for this investigation.

Using logarithmic axes the data for tones shorter than 50 cycles is approximated well by a straight line of gradient −1. This indicates that the relative FDL ($\Delta f/f$) is inversely proportional to the number of cycles ($\mathcal{N}$) presented. The physical fractional bandwidth (BW/f) of a pure tone is also inversely proportional to the number of cycles ($\mathcal{N}$) of the signal. This suggests that the threshold is limited by the shape of the spectral envelope of the tone pulses (Moore[6]). We can write an uncertainty[2] relationship:

$$\Delta f/f = \beta/\mathcal{N} \tag{1}$$

where β is a constant.(From Fig.1 $\beta \approx 0.08 \pm 0.02$)

The relative FDL data does not vary significantly with $\mathcal{N}$ above 100 cycles. This suggests that the spectral width of the tone pulses is no longer limiting the frequency resolution and the 'long-duration' asymptote of the threshold has been reached.

2.2.3 Pitch as Log(Frequency). Fig.1 also indicates that the effect of frequency is largely removed by using the relative FDL ($\Delta f/f$). This section aims to show that the relative FDL can be thought of as a *Pitch* Difference Limen, which would be expected to be independent of frequency. Note that it is not intended to validate a new definition for the term pitch which already has too many uses.

[1] Freyman & Nelson[3] used three straight lines to fit FDL data for normal-hearing subjects. This was for the purposes of testing a variation of Zwicker's excitation-pattern model of frequency discrimination.

[2] Moore[7, p131–135] explains the use of Signal Detection Theory to show that a psychoacoustic threshold can be related to the uncertainty (variance) of the related 'internal' variable.

Pitch is defined as "*that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale*" (American Standards Association, 1960). This musical scale is clearly not a linear function of frequency, *e.g.* the difference between 100 & 200Hz does not sound the same as that between 200 & 300Hz. However, a musical melody composed of notes which have frequencies at fixed ratios to each other sounds similar if the reference frequency is changed. This suggests the idea of using a logarithmic mapping from frequency to *Pitch*[3].

$\mathcal{P} \propto \log f$

From this the threshold detectable difference in *Pitch* between two tones — the Pitch Difference Limen (PDL), $\Delta\mathcal{P}$, can be defined as:

$$\begin{aligned} \Delta\mathcal{P} &= \mathcal{P}_2 - \mathcal{P}_1 \\ \Rightarrow \Delta\mathcal{P} &\propto \log(f_2/f_1) \end{aligned} \quad (2)$$

$$\Rightarrow \Delta\mathcal{P} \propto \log(1 + \frac{\Delta f}{f_1}) \quad (3)$$

For $\Delta f/f \ll 1$ this approximates to:

$$\Delta\mathcal{P} \propto \Delta f/f \quad (4)$$

So the relative FDL is equivalent to a PDL independent of frequency, and the uncertainty relationship of equation (1) for $\mathcal{N} < 50$ becomes:

$$\Delta\mathcal{P} \propto 1/\mathcal{N} \quad (5)$$

3. TRANSIENT PITCH SHIFT DETECTION

In most PDL experiments two consecutive, stationary tones slightly different in frequency are played. If these tones are made contiguous a dynamic stimulus containing a transient pitch shift is formed. The uncertainty relationship equation (5) applied to this simple argument, indicates that short-duration transient pitch shifts should be less audible. To confirm this for dynamic tones, Frequency Modulation Detection Thresholds[4] (FMDT), Δf_c, need to be investigated.

3.1 Previous work

As pitch transients are most likely to occur at the start of a sound, followed by a more stable section of tone, emphasis was placed on data for stimuli of this form. Many authors results (*e.g.*, Sergeant & Harris[9]; Pollack[8]) showed that the relative FMDT ($\Delta f_c/f_c$) for frequency glides varied with $\mathcal{N}$ (number of cycles of the carrier frequency) in a similar way to the PDL for stationary tones. Relative FMDT results from Carlyon & Stubbs[1] for single-cycle sinusoidal F.M. of pure tones were also comparable to PDLs. This suggests a sampling model for frequency modulation detection, in which listeners sample the stimuli as a means of detecting the change in frequency.

Dooley & Moore[2] examined frequency discrimination thresholds for both stationary and gliding tones. They compared the data obtained with the predictions of a simple sampling model and concluded that the simple theory was consistent with the data.

[3] Thus (all) the notes of a piano keyboard would be equally spaced on this *Pitch* scale

[4] The FMDT is the size of carrier frequency shift corresponding to the threshold for frequency modulation detection. It will be used to refer to the threshold for all types of frequency modulation. (*e.g.* glides, sinusoidal modulation, etc...)

ON THE AUDIBILITY OF TRANSIENT PITCH SHIFTS

Horst[4] used harmonic complex, dynamic stimuli with initial and final fundamental frequencies the same to discourage discrimination on the basis of differences in these frequencies, (Tsumura *et al.*[10] suggested that the dominant cue for glide detection was the pitch difference between the initial and final portions of the stimulus). The report indicates that a sampling model predicts the right order of magnitude of FMDT for such stimuli. Horst[4] and Carlyon & Stubbs[1] have both shown that FMDTs are reduced by introducing more harmonics of the fundamental frequency.

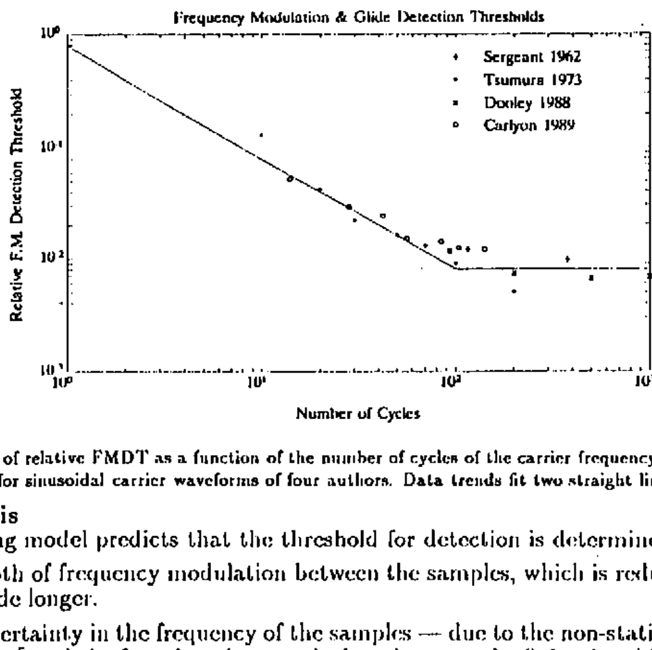

Figure 2: Plot of relative FMDT as a function of the number of cycles of the carrier frequency ($\mathcal{N}$) (log scale on both axes). Results for sinusoidal carrier waveforms of four authors. Data trends fit two straight lines as shown.

3.2 Analysis

The sampling model predicts that the threshold for detection is determined by optimising:

- the depth of frequency modulation between the samples, which is reduced if the samples are made longer.
- the uncertainty in the frequency of the samples — due to the non-stationary pitch during sampling[5] and the fact that the sample duration must be finite, in which case the sample accuracy must be constrained by the stationary tone PDL.

Fig.2, of data normalised as in section 2.2.1, indicates that for short-duration modulations ($\mathcal{N} < 100$) the threshold has an inverse relationship to $\mathcal{N}$ as for the PDL, but is about 10 times as large. This indicates that for short-duration modulations the most significant uncertainty in the sample frequency is that due to the finite sample duration. As there must be more than one sample during the modulation in order to detect it, the sample duration must be significantly smaller than the overall duration of the stimulus. Thus it is to be expected that the relative FMDT corresponds to a PDL for a smaller value of $\mathcal{N}$.

For modulations with $\mathcal{N} > 100$ the relative FMDT is almost constant, indicating that the sample accuracy is no longer the limiting factor.

[5] Dooley & Moore[2] ignored this contribution to the uncertainty.

4. EXPERIMENT

4.1 Introduction

The main aim of this experiment was to test the hypotheses formulated, with stimuli and under conditions more closely resembling those encountered by audio systems.

Untrained subjects were used in order to establish the thresholds for 'normal' conditions, in which listeners are not highly trained at detecting a very specific event. Under normal circumstances listeners would receive a wide variety of different stimuli, which would prevent them from developing acuity to changes in a particular parameter. To prevent any improvements due to learning from systematically affecting the results, the order in which each subject did the tests was randomised.

In music a wide range of notes are usually played and so listeners do not expect each successive tone to be at the same frequency. Thus the tones used were randomised in frequency by a small amount. Also the output from an electronic synthesiser would contain a range of harmonic structures, not just pure tones, thus harmonic complex tones were used.

4.2 Method

4.2.1 Overview. A two-interval, two-alternative, forced-choice task was used. Subjects were presented with pairs of tones, one stationary in frequency and one starting with a downward linear frequency glide. The subjects' task was to identify the tone starting with a frequency transient. The two tones of equal amplitude were separated by a silent interval of 200ms. The tone pairs were spaced 3s apart and grouped in blocks of ten pairs, each block spaced by 10s. The frequency transient was followed by approximately 100 cycles of constant frequency in all cases. This section of signal had the same fundamental frequency as the stationary tone. The frequency transient duration, added to the time for 100 cycles at the fundamental frequency, determined the duration of both tones in a pair.

Each test sequence, lasting less than ten minutes, consisted of tones of one fundamental frequency and one duration only. Signal fundamental frequencies of 100, 400 and 1000Hz combined with frequency transient durations of 50, 100 and 300ms were used. Thus there were a total of nine test sequences. To prevent the subjects from memorising the reference frequency and responding to any stimulus containing energy at other frequencies, every tone's fundamental frequency was randomised by 3%. A typical trial is shown in Fig. 3.

4.2.2 Stimuli. All signals were harmonic complexes, consisting of the first five harmonic components of the fundamental frequency, at the same amplitude and initial phase. The data was generated on an Atari 1040ST computer, using the "Composers Desktop Project", and directly recorded onto Digital Audio Tape (DAT). The sampling frequency used was 44100Hz and the frequency and amplitude parameters were updated every ten sample periods. All signals started at positive going axis crossings and used one cycle rise and fall 'times' to minimise switching transients at the start and end of the signals. The signals were delivered to subjects diotically using Beyer Dynamic closed-back DT100 headphones.

4.2.3 Procedure. A non-adaptive, two-interval, two-alternative, forced-choice procedure was used. The frequency transient occured with equal probability in one of the two intervals of a trial. For each condition (*i.e.* combination of fundamental frequency and transient duration) five values of transient depth, δf_0 were used. These values were chosen following

ON THE AUDIBILITY OF TRANSIENT PITCH SHIFTS

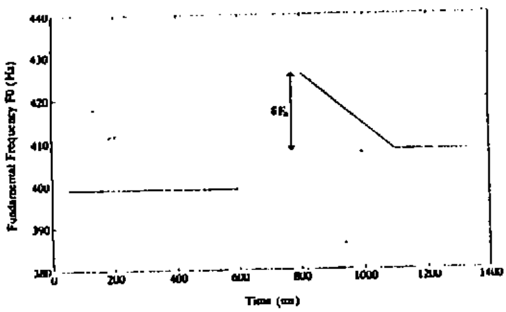

Figure 3: A schematic representation of the test tones used. The two stimuli in a trial occured at random in the order shown or in the reverse order. The frequency transient has depth δf_0, measured as the change in the fundamental frequency f_0, and is followed by 100 cycles of stationary tone.

informal listening tests by the authors, to range from easily audible and to encompass the 75% correct response level. Each value of δf_0 was used twenty times within the test sequence, giving a total of 100 trials per sequence which were randomised in order.

The 75% correct response level, which is generally taken as the threshold value, was estimated for each condition as follows:

1. 95% confidence limits were calculated for the correct response level (%CR) corresponding to each transient depth.
2. %CR was plotted against δf_0 on a 'probability' graph, (with this non-linear scale a 'normally' distributed parameter produces a straight line).
3. Δf corresponding to 75%CR was determined from a best fit line through the data points.
4. An estimate of the uncertainty of this value was made by considering a number of possible straight lines that would pass within the 95% confidence limits.

4.2.4 Subjects. Three subjects (CB, MP and PS) participated in the experiment and were tested with all nine sequences. Subjects were not trained in frequency modulation detection tasks. They were allowed to practice with a demonstration sequence similar to the test sequences, but containing some more easily audible frequency glides, until they understood the nature of the task. The subjects listened to the nine test sequences over a period of a few days, never performing more than one ten minute session per hour.

4.3 Results and Discussion

As the results for each subject were not significantly different from each other the data was averaged before calculating the %CR and subsequent Δf values.

4.3.1 Comparison to previous work. Fig.4 shows the FMDT data obtained in this experiment and the best fit lines to results by other authors from section 3.2.

For $\mathcal{N}$ small the relative FMDT results are in agreement with those of other authors, whom used trained subjects (*e.g.*, Tsumura[10]; Carlyon & Stubbs[1]), following an inverse proportionality to $\mathcal{N}$. This suggests that for small values of $\mathcal{N}$ the FMDT is limited by a fundamental

ON THE AUDIBILITY OF TRANSIENT PITCH SHIFTS

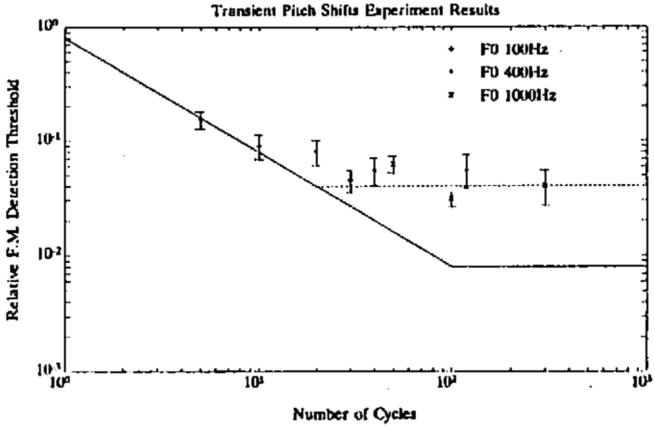

Figure 4: Plot of the relative depth of the Frequency Transient ($\Delta f/f$) at the 75% correct response level as a function of the Transient duration as numbe of cycles of the fundamental frequency ($\mathcal{N}$). Data for average of the three subjects. Dashed line shows approximate fit to data for $\mathcal{N} > 20$. Solid line shows the best fit line from Fig.2.

constraint for all listeners. For $\mathcal{N} > 20$ the relative FMDT is approximatly constant at 4%. Clearly the lack of training prevented listeners form attaining the minimum threshold.

4.3.2 Transient detection under 'real' conditions. In most situations music is not listened to over headphones. While performing the experiment it was observed (and confirmed by testing) that the frequency transients were much easier to detect for trials using a loudspeaker in a normal room instead of headphones. A similar decrease in threshold was observed if the test tones were digitally processed to add typical room echos and reverberation (by convolution with a room impulse response) and then presented over headphones. This confirmed that the effect was at least in part due to the complex frequency response of a typical room.

Thus under 'real' conditions the frequency transient is accompanied by potentially large fluctuations in level which are not present for the stationary tone, thus providing additional loudness cues for identification of the target tone. For harmonically rich tones each component experiences different changes in amplitude as the fundamental frequency shifts, thus there is also a variation in the timbre of the tone.

For applications involving pitch extraction for control of electronic music it is only important that the listener does not perceive a change in pitch (as amplitude and timbre variations will occur naturally as part of most synthesiser sounds). Thus the lower thresholds obtained with loudspeaker trials, in which the listener was not hearing the transient pitch shift, *per se*, but used other cues to detect the target tone, are not necessary. However, in some applications, where it is undesirable for any changes to be noticable, lower thresholds would be required, dependant also on the detectability of amplitude modulation.

ON THE AUDIBILITY OF TRANSIENT PITCH SHIFTS

5. CONCLUSIONS

This paper has shown that it is the number of cycles of a stimulus, rather than the absolute duration, that is important for frequency differency and frequency modulation detection thresholds. Through the use of this and *Pitch*, regarded as log frequency, results from several previous authors have been combined. This data indicates a pitch uncertainty realationship for stationary tones, inversely proportional to the stimulus number of cycles and provides support for a sampling model of frequency modulation detection. Further, the experiment has shown that subject training and the use of headphones in psychoacoustic threshold experiments may lead to specifications that are too stringent or too lax respectively for particular acoustic applications. Further experiments with both loudspeakers and headphones, for different shapes of pitch transient are required. The 'fundamental' nature of the FMDT for very short-duration modulations has also been highlighted.

References

[1] Carlyon, R. P. & Stubbs, R. J. (1989). *"Detecting single-cycle frequency modulation imposed on sinusoidal, harmonic & inharmonic carriers,"* J. Acoust. Soc. Am. **85**, 2563–2574.

[2] Dooley, G. J., & Moore, B. C. J. (1988). *"Detection of linear frequency glides as a function of frequency and duration,"* J. Acoust. Soc. Am. **84**, 2045–2057.

[3] Freyman, R. L. & Nelson D. A. (1986). *"Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners,"* J. Acoust. Soc. Am. **79**, 1034–1044.

[4] Horst, J. W. (1989). *"Detection and discrimination of frequency modulation of complex signals,"* J. Acoust. Soc. Am. **85**, 2022–2030.

[5] Mark, H. E. (1990). *"Frequency discrimination of single-, double-, and triple-cycle sinusoidal acoustic signals,"* J. Acoust. Soc. Am. **88**, 560–563.

[6] Moore, B. C. J. (1973). *"Frequency difference limens for short-duration tones,"* J. Acoust. Soc. Am. **54**, 610–619.

[7] Moore, B. C. J. (1980). *"An Introduction to the Psychology of Hearing, 3rd. Ed.,"* (Academic Press).

[8] Pollack, I. (1968). *"Detection of rate of change of auditory frequency,"* J. Exp. Psychol. **77**, 535–541.

[9] Sergeant, R. L. & Harris, J. D. (1962). *"Sensitivity to unidirectional frequency modulation,"* J. Acoust. Soc. Am. **34**, 1625–1628.

[10] Tsumura, T., Sone, T., & Nimura, T. (1973). *"Auditory detection of frequency transition,"* J. Acoust. Soc. Am. **53**, 17–25.

[11] Turnbull, W. W. (1944). *"Pitch discrimination as a function of tonal duration,"* J. Exp. Psychol. **34**, 302–316.