DAT335 – Music Perception and Cognition
Cogswell Polytechnical College
Spring 2009

Week 5 – Class Notes

Temporal Processing in the Auditory System

Introduction

Time is an important dimension in hearing, since almost all sounds fluctuate over time. Such sounds,like music and speech, can convey most of their information in the fluctuations themselves, rather than in the parts of the sounds which are relatively stable. Temporal event detection can be classified into two categories: temporal resolution (acuity), or the ability to detect changes in stimuli over time; and temporal integration (summation), of the ability to add up information over time to enhance the detection or discrimination of stimuli.

Temporal resolution can be characterized by the filtering that takes place in the peripheral auditory system. Temporal resolution depends on analysis of the time pattern occurring within each frequency channel, and the comparison of the time patterns across channels.

A difficulty encountered when measuring temporal resolution of the auditory system is that the changes in the time pattern of a sound are generally associated with with changes of its spectral magnitude. Thus, temporal detection depend not only on the the temporal resolution, but on the detection of spectral change as well.

Temporal Resolution Measured by the Discrimination of Stimuli with Identical Magnitude Spectra: Broadband Sounds

The Detection of Gaps in Broadband Noise

The threshold for detecting a gap in a broadband noise provides a simple and efficient measure of temporal resolution. A 2AFC procedure is used and the subject is to indicate from a series of two noise burst which one presents an interruption (at random). The measured gap threshold lies around 2-3 ms. The threshold increases at very low sound levels, while it remains rather invariant for moderate to high levels.

The Discrimination of Time-Reversed Signals

The long-term magnitude spectrum of a sound does not change when the sound is time reversed (played backward). If a time-reversed sound can be discriminated from the original, then this must reflect a sensitivity to the difference in time pattern of the two sounds. Studies show that subjects could distinguish gaps in the presented pattern of 2-3 ms. Further experiments, suggest a systematic variation in threshold as a function of both overall sound level and the relative level of the individual stimuli (in this case clicks). The resulting thresholds ranged from 0.5-1.8 ms.

Temporal Modulation Transfer Functions

This approach measures the threshold for the detection of changes in the amplitude of a sound as a function of the rapidity of the change. The function relating threshold to modulation rate is known as the temporal modulation transfer function (TMTF). In the simplest case, white noise is sinusoidally amplitude modulated, and the threshold for detecting modulation is determined as a function of the modulation rate. The modulation of white noise does not change its long-term magnitude spectrum.

For low modulation rates, performance is limited by the amplitude resolution of the ear and not by its temporal resolution. The threshold in this case is independent of the modulation rate for rates up to 16 Hz. Rates beyond this frequency and up to 1000 Hz, do show an effect in the temporal resolution. Above 1000 Hz the modulation cannot be detected.

Sensitivity to amplitude modulation decreases progressively as the rate of modulation increases.

Temporal Resolution Measured by the Discrimination of Stimuli with Identical Magnitude Spectra: Effects of Center Frequency

The previously mentioned experiments use broadband stimuli, and as such do not provide any information regarding the question of whatever the temporal resolution of the ear varies with center frequency. It is often suggested, in theory, that temporal resolution should be quite poor at low frequencies than at higher ones. This idea is supported by the response time of the auditory filters. For example, is a stimulus is turned off and on again to form a temporal gap, ringing in the auditory filters would partially fill in the gap, so that the envelope of the auditory filter output only shows a small dip. The narrower the filter, the longer its response time. Since auditory filters have narrower bandwidths at low frequencies than at high, then their responses limit the temporal resolution. Thus, resolution of auditory filters at low frequencies is worse than those at higher frequencies.

Measured thresholds of 1-2 ms were observed for center frequencies at 2-4 kHz. However, for a center frequency of 1 kHz, the threshold was slightly higher (2-4 kHz). These results suggest that the response of auditory filters is not important for filters above 2 kHz, but they might play a role below that.

Measurement of Temporal Resolution Using Narrowband Sounds

For narrowband sounds, changing the time pattern changes the spectrum as well. If, for example, a gap is introduced to such a sound; the spectrum will be splattered (the energy is spread outside the nominal bandwidth of the sound. Spectral splatter detection is prevented by masking the sound with a noise background.

Detection of Gaps in bands of Noise

Gap thresholds for noise bands may depend on two factors: the inherent fluctuations in noise and the effects of the auditory filters. For example, consider a sample of noise with a gap that has been passed through an auditory filter (the filter is centered at the passband of the noise). The rapidity of random fluctuations in the noise increases with increasing bandwidth. It has been suggested that gap thresholds for noise band may be limited by the inherent fluctuations in the noise. Thus, randomly occurring dips in the noise may be "confused" with the gap to be detected.

If the noise bandwidth is less than that of the auditory filter, the temporal pattern of the fluctuation in the noise is hardly changed by the auditory filter. The filter's ringing partially "fills" the gap, but not completely. The gap threshold the depends primarily on the confusability of the gap with the inherent fluctuations of the noise itself. Narrow bandwidths lead to slower fluctuations, which in turn lead to larger gap thresholds. The gap threshold is thus expected to decrease with increasing bandwidth.

On the other hand, if the noise bandwidth is greater than the auditory filter bandwidth, the fluctuations at the filter's output are slower than those for the input. To reliably detect a gap by monitoring the filter's output, the gap has to be longer than a typical dip in the output of the filter. In this case, the filter's bandwidth is the limiting factor rather than the input noise bandwidth. Thus, if subjects monitored just a single auditory filter, the effect of the filter would impair the gap detection.

Based on the previous statements, it could be predicted that (if a single auditory filter is to be used) the detection of temporal gaps in noise bands should be worse for low center frequencies than at high center frequencies. This is because the filter's bandwidth are smaller at lower center frequencies, thus leading to slower fluctuations and larger gap thresholds. Additionally, gap thresholds should decrease with increasing noise bandwidth, but only if the noise bandwidth is less than that of the widest auditory filter stimulated by the noise.

However, this expected pattern is not shown in empirical data. Gap thresholds for narrowband noise do decrease with increasing bandwidth, but they will continue decreasing even for noise bandwidths exceeding the bandwidth of any single stimulated auditory filter.

It seems that subjects use multiple auditory filters to detect temporal gaps in broadband noises. When a gap is introduced into a noise band with a large width, the gap creates a dip that is synchronous at the outputs of all the auditory filters that are excited by the stimuli. Because the randomly occurring dips in the noise are independent at the auditory filters output, better gap detection is achieved by using multiple filters (as opposed to using a single filter). As a result, the gap detection threshold decrease (improves) with increasing noise bandwidth, even when the bandwidth greatly exceeds the bandwidth of any single auditory filter that is excited by the noise. Combining the information from multiple filters do not limit gap detection, and the center and upper cutoff frequencies do not have a significant effect.

Detection of Gaps in Sinusoids

In studies regarding the detection of gaps in sinusoids, the result were strongly affected by the phase at which the sinusoid was turned off and on to produce the gaps. A noise background with a notch at the sinusoids frequency in order to mask the spectral splatter. Three phase conditions (phase difference in which the waveform was turned on after the gap) were used. The "standard" phase has a sinusoid starting at a positive-going zero-crossing; the "reversed" phase condition has the sinusoid starting at a negative-going zero-crossing; and for the "preserved" phase condition,the sinusoids starting phase is the same as that if it had continued. The preserved-phase condition the gap has been "cut off" from a continuous sinusoid.

For a 2AFC test of a sinusoid at a given frequency, shows that the preserved-phase condition improves with increasing gap duration. For the other two conditions, gap detection is difficult if its value is an integer multiple of the period (P). The gap is easily detected for standard-phase conditions if the gap has a length of $(n + 0.5)P$, where $n = 0,1$. Reversed-phase condition is poorly detected for gaps of duration $(n + 0.5)P$, where $n = 0,1$. In the gap duration is nP the detection if better.

These results can be explained in terms of the ringing in the auditory filter. If a sinusoid is turned off at the start of the gap, the filter will continue to respond or ring for a certain time. If the gap duration corresponds to one whole period of the sinusoid, the sinusoid following the gap is in phase with the ringing response. In this case, the filters output shows a small dip, and as a result gap detection is difficult. For gap durations that are not in phase with the ringing response, the filters output shows a zero before returning to a steady-state. The resulting dip is larger and easier to detect.

It appears that the ringing of an auditory filter only limits gap detection for sinusoidal waves at very low frequencies.

Temporal Modulation Transfer Functions for Sinusoidal Carriers

TMFT's for higher carrier frequencies generally show an initial flat portion (where sensitivity is independent of modulation frequency), then a portion where the threshold increases with increasing modulation frequency (presumably reflecting the limits of temporal resolution), and then a portion where the threshold decreases again (presumably reflecting the detection of spectral sidebands).

The initial flat portion of the TMFT suggests a discrepancy of the inhering amplitude fluctuations in a noise carrier, which limit the detectability of the imposed modulation. Narrowband noise carriers show an inherently slow amplitude modulation, which in turn make the TMFT show a poor sensitivity for low modulation frequencies. Thus, TMFT's obtained with narrowband noise is a poor measure of temporal resolution. For this reason, TMFT's measured using sinusoidal carriers are often preferred.

Modeling Temporal Resolution

While there is evidence that the auditory filter plays a role in some measures of temporal resolution, its influence is seen mainly at low frequencies. The auditory filter's response at high center frequencies is too fast to be a limiting factor in most tasks involving temporal resolution. As a result, it has been though that there is a process at levels of the auditory system higher than the auditory nerve which is considered “sluggish” in some way, therefore limiting temporal resolution. Models of temporal resolution, concerned with this process, assume that the internal representation of a stimuli is “smoothed” over time; rapid temporal changes are reduced in magnitude, but slower ones are preserved. This smoothing process almost certainly operates on neural activity, however the most widely used models are based on smoothing out a simple transformation of the stimulus., rather than its neural representation. This particular model although it is not very realistic, it is rather simple mathematically.

Bandpass Filtering

This initial bandpass filtering stage represents the action of the auditory filter. For simplicity's sake only one filter is shown, but in reality there is an array of parallel channels with its own bandpass filter.

The Nonlinearity

Following the filter there is a nonlinear device. This nonlinearity is meant to reflect the operation of several processes occurring in the peripheral auditory system. For example, nerve spikes occurring at specific phases of the stimulating waveform on the BM, or the compressive input-output function of the BM. Recent models of temporal resolution include these two processes: rectification and compression.

The Sliding Temporal Integrator

The output of the nonlinear device is fed into a “smoothing” device, which can be implemented as a lowpass filter or a sliding temporal integrator. This device determines a weighted average of the output of the compressive nonlinearity over a certain time inerval or “window”. This function is often modeled as a pair of back-to-back exponential function and is called the “shape” of the temporal window. Most weight is given to the nonlinear device's output at times close to the temporal center of the window, and progressively less weight to the output at time farther from the center. The window itself slides in time, so that the output of the integrator is like a running average of the input. This process has the effect of smoothing rapid fluctuations while preserving slower ones. Rapidly occurring sounds take some time to either build up or decay down.

It is assumed that both backward and forward masking depend on the process of build up and decay. For example, if a brief signal is rapidly followed by a masker, the response to the signal may still be building up when the masker occurs. If the intensity of the masker is large, then its effects may “swamp” those of the signal. Similarly, for a brief signal that is followed by a masker, the decaying response to the masker may swamp the response to the signal.

The Decision Device

The output of the temporal integrator is fed into a decision device. This device may use different “rules” depending on the task at hand. For example, if the task is to detect a brief temporal gap in a signal, the decision device will look for “dips” in the output of the temporal integrator. If the task is to detect amplitude modulation of a sound, the device might assess the amount of modulation at the output of the sliding temporal integrator.

A Modulation Filter Bank?

Researchers have suggested that the analysis of sounds that are amplitude modulated depends on a specialized part of the brain that contains an array of neurons, each tuned to a different modulation rate. As such, each neuron can be considered a filter in the modulation domain, and the array of neurons is knows as a “modulation filter bank”. Neurons with appropriate properties have been found in the cochlear nucleus. However, this remains a rather controversial subject.

Masking in the Modulation Domain with Broadband Carriers

Modulation masking refers to the increasing amplitude modulation detection threshold of a given carrier if additional amplitude modulation is superimposed on that carrier. Studies in the detection of sinusoidal amplitude modulation of a carrier (for example pink noise) involve the measurement of the modulation threshold detection when no other modulation is present and when a “masker” modulator is added. For each masker, the masking pattern shows a peak in the masker frequency. These results are interpreted as a selectivity in the modulation-frequency domain.

TMFT's Obtained Using Narrowband Carriers

Result from these experiments show that the detection of the signal modulation was more difficult when its frequency fell within range of the random amplitude modulation inherent in the carriers. Also, the sharpness of the tuning of the hypothetical modulation filter bank is much less than the sharpness of tuning of the auditory filters in the frequency domain. The modulation filters, if they exist, are not very selective.

Modulation Detection Interference

The detection of amplitude modulation of a sinusoidal carrier can be impaired by the presence of one or more modulated sounds with different carrier frequencies. This effect is known as “modulation detection interference” (MDI) or “modulation discrimination interference”.

For example, one particular study shows that the threshold for detecting sinusoidal amplitude modulation of a sinusoidal carrier was increased in the presence of another carrier amplitude modulated at the same rate, even when the second carrier was remote in frequency from the first. It was also found that MDI did not occur if the second carrier was unmodulated. This effect seems to show a tuning for modulation rate, being greatest when the modulation rates of the target and interfering sounds were close.

It has been suggested that MDI occurs because the target and interferer modulation are processed together in a modulation filter bank, where each filter is tuned for a modulation rate, but responds to a broad range of carrier frequencies. However, this concept does not fit well with the physiological data. These show that most neurons with tuning in the modulation-frequency domain are also tuned to the carrier-frequency domain.

Accounting for TMFT's with a Modulation Filter Bank

An alternative approach for the shape of a TMFT for broadband noise carriers lies in the use of a modulation filter bank. Modulation at a specific frequency is assumed to be detected by monitoring the output of a modulation filter tuned close to that frequency. The ability to detect modulation is assumed to be partly limited by the inherent random amplitude fluctuations in the noise as they appear on the modulation filter's output. The bandwidths of the modulation filters increase with increasing center frequency. Simply put, the more random the modulation appears at the output of the modulation filters tuned to higher frequencies, the more difficult it is to detect modulation as the modulation frequency increases.

Duration Discrimination

Studies implemented to detect a change in the duration of an auditory stimuli usually involve two successive sounds that have the same power-spectrum, but different duration. The subject is required to indicate which sound has the larger duration. The smallest detectable increase in duration is labeled ΔT, while the baseline duration is T. Results show that the Weber fraction, $\Delta T/T$, decreased with increasing T.

All studies show that, for values of T exceeding 10ms, ΔT increases with T and ΔT is roughly independent of the spectral characteristics of the sound. However, ΔT increases at low sound levels.

Temporal Analysis Based on Across-Channel Processes

Subjects in studies of the ability to compare timing across different frequency channels may be able to distinguish different time patterns without the subjective impression of a change in time patters; rather, a change in the quality of the sound is heard.

The Discrimination of Huffman Sequences

Huffman sequences describe a class of signals that have the same long-term magnitude spectrum, but different short-term spectrum. Essentially, they are brief broadband sounds except that their energy content in a certain region is delayed relative to that of other regions. The amount of delay, center frequency of the delayed region, and the width of the delayed region can be varied.

If subjects can distinguish a pair of Huffman sequences differing, for example, in the amount of delay in a given region, it implies that they are sensitive to the difference in time patterns. Some studies show that subjects could detect time differences in delay times of about 2ms regardless of the center frequency of the delayed region. However, the temporal pattern differences were perceived as subtle changes in sound quality and that extensive training of a subject is required to achieve a fine acuity.

Detection of Onset and Offset Asynchrony in Multicomponent Complexes

One study measured the thresholds for detecting asynchrony in the onset or offset of complex signals composed of many sinusoidal components. These components were either uniformly spaced on a logarithmic frequency scale or formed a harmonic series. In one stimulus, the standard, all components are started and stopped synchronously. In the “signal” stimulus, one component is presented with an onset of offset asynchrony. The subjects task is to discriminate the standard stimulus from the signal stimulus.

Results of this test show that onset asynchrony was easier to detect than offset asynchrony. For harmonic signals onset asynchronies os less than 1ms could be detected. Thresholds for detecting offset asynchronies were larger, 3-10ms when the asynchronous component ended after the other components and 10-30ms when the asynchronous component ended before the other components. Overall, thresholds for detecting asynchronies in logarithmically spaced complexes were 2-50 times larger than for harmonic complexes.

This difference may lie within the concept of perceptual grouping. Harmonic signals are perceived as a single sound source, while logarithmically spaced complex sounds were perceived as a series of separate tones. The high sensitivity to onset asynchronies for harmonic complexes is consistent with the finding that the perceived timbres of musical tones are partly dependent on the exact onset times and rates of rise of individual harmonics within each musical note.

Judgement of Temporal Order

The ability to judge the temporal order of a sequence of sounds strongly depends on whether the task requires identification of the order of individual elements or whether It can be performed by discrimination of different orders or by attaching well-learned labels to different orders. For the latter case, resolution can be rather fine, 10ms or less, if extended learning and feedback is done. For tone sequences, the component durations necessary for labeling different orders can be as low as 2-7ms. This type of acuity is expected for speech sounds, since they consist of well-learned sequences to which consistent labels have been attached.

However, when the task is to identify the order of sounds in a sequence, performance is rather poor. In general, when the number of items in the sequence is increased, so do the durations required for order identification.