



Sensitivity to temporal fine structure probed with pulse-spreading harmonic complexes

Olivier Macherey

► To cite this version:

Olivier Macherey. Sensitivity to temporal fine structure probed with pulse-spreading harmonic complexes. International Symposium on hearing, Jun 2022, Lyon, France. hal-03876570

HAL Id: hal-03876570

<https://cnrs.hal.science/hal-03876570>

Submitted on 28 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sensitivity to temporal fine structure probed with pulse-spreading harmonic complexes

Olivier Macherey^{1,*}

¹ Aix-Marseille University, CNRS, Centrale Marseille, LMA, 4 Impasse Nikola Tesla, 13013 Marseille, FRANCE

* Contact author: macherey@lma.cnrs-mrs.fr

Abstract

Temporal fine structure (TFS) is assumed to play an important role in both pitch and speech processing in normal hearing (NH). However, there is still some debate on how TFS is coded in the auditory periphery. Moore & Sek (2009) provided evidence that NH subjects are sensitive to TFS at very high frequencies in stimulus conditions where spectral cues should not be present. This is important because it implies that phase locking may still provide useful information above 4000 Hz where it is commonly believed to be absent. Here, we examine sensitivity to TFS using different stimuli than the frequency-shifted complexes previously used. Pulse-spreading harmonic complexes (PSHCs) are spectrally-dense complexes with a low f_0 (here 2 Hz). The phase of the components is set so that the envelope rate can be manipulated independently from the f_0 . Here, we use a special case of PSHCs for which the TFS peaks were shifted relative to the envelope peaks from one envelope period to the next. For upward PSHCs (up-PSHC), the TFS peaks were progressively advanced, yielding a rising pitch percept across the stimulus, while for downward PSHCs (down-PSHC), the peaks were progressively delayed, yielding a falling pitch percept. In Experiment 1, subjects heard up-PSHC, down-PSHC and regular inharmonic complexes bandpass-filtered with a lower cut-off (F_c) of 2000 Hz and were asked to identify whether the pitch was rising, falling or flat in a one interval, three alternative forced choice task. The envelope rate was 98, 200 or 450 pps. For the two highest rates, most subjects correctly identified the direction of the pitch change while for the lowest rate, the stimuli could not be discriminated. This shows that, for a given F_c , there is an envelope rate below which the auditory system is insensitive to changes in TFS. The lower rate limit of TFS sensitivity was measured in Experiment 2 using an adaptive three interval, two alternative forced choice task where subjects heard two down-PSHC and one up-PSHC and had to pick the latter. The stimuli were bandpass filtered in seven frequency regions with F_c varying from 250 to 11200 Hz and were presented in the presence of threshold-equalising noise. The lower limit significantly increased from 42 pps to 691 pps with increases in F_c . The ratio between F_c and the envelope rate at threshold also increased from 6 to 18 when F_c increased from 250 to 2000 Hz, then remained flat with further increases in F_c or slightly decreased at the highest F_c . In the lower part of the spectrum (<2000 Hz), this lower limit resembles the lower limit of melodic pitch. The results in the upper part of the spectrum suggest that TFS cues are still available in high frequency regions (up to 10 kHz) when the rank of the lowest harmonic present in the passband is 18, and all harmonics are presumably unresolved.

Introduction

The spectral decomposition performed by the cochlea resembles that achieved by a bank of overlapping bandpass filters: in response to a sound, the vibration pattern of a particular point on the basilar membrane corresponds to a bandpass-filtered version of the original sound waveform. This filtered waveform can be decomposed into a slowly varying envelope superimposed on a carrier, referred to as the temporal fine structure (TFS), which oscillates at a rate close to the characteristic frequency of this particular cochlear location. Auditory nerve fibers can, to some extent, phase-lock to this TFS and there is evidence that this temporal neural information is used by the binaural system for frequencies up to about 1400 Hz (Zwislocki & Feldman, 1956). The exact limit of phase locking in humans and whether or not it is important for pitch perception remain, however, a matter of debate. Most animal species exhibit a steep decrease in phase locking for frequencies above about 1000 or 2000 Hz (Koppl, 1997). The finding that pure tone frequency discrimination in humans also deteriorates for frequencies above 2000 Hz has often been cited as evidence that frequency discrimination depends on phase locking information (e.g., Moore & Ernst, 2012). However, Oxenham et al. (2011) reported that harmonic complex tones only containing components above 4000 Hz can still convey a clear, musical pitch percept. They concluded either that phase locking information is not needed for complex pitch perception or that it is still usable by humans at these very high frequencies.

Moore & Sek (2009) investigated sensitivity to TFS at very high frequencies (above 8 kHz) by asking their subjects to discriminate between bandpass filtered harmonic and inharmonic complex tones. In each trial, the inharmonic complex contained the same components as the harmonic complex except the components were shifted by a fixed amount in Hz. These two stimulus waveforms have identical envelopes but different TFS. Moore & Sek (2009) showed that subjects could discriminate between these two stimuli when the lowest harmonic present in the passband was the 12th. According to usual definitions of resolvability (Moore & Gockel, 2011), this would imply that the components were unresolved and that the subjects could not use differences in excitation patterns to perform the task. They, therefore, concluded that TFS sensitivity was mediated by phase locking information which in turn implied that phase locking cues could still be used by their subjects at frequencies higher than 8 kHz. This is controversial because as previously mentioned, animal experiments show poor phase locking at these frequencies and there are currently no human single-unit data available in the literature. Note, however, that significant phase locking up to 7 kHz in chinchillas were obtained by Kale and Heinz when averaging a lot of data to lower the noise floor of their recordings, thereby suggesting there may still be usable temporal information even at these high frequencies (Verschooten et al., 2019). A recent opinion paper on the use of phase locking cues in humans illustrates this controversy: estimates of the highest frequency at which phase-locking information could be used by humans ranged across “experts” in this field from 1.5 kHz to 10 kHz (Verschooten et al., 2019).

One draw-back of using frequency-shifted inharmonic complex tones to investigate TFS sensitivity is that these stimuli produce ambiguous pitch percepts and that subjects were not explicitly asked to compare the pitches of inharmonic and harmonic complex tones; any audible difference between the harmonic and inharmonic tones could be used as a cue. Furthermore, harmonic and inharmonic complexes produce different envelopes at the output of a given auditory filter which can potentially provide another cue to discriminate them. This led Moore & Sek (2009) to randomize the starting phase of their components, which in turn affected the shape of the stimulus envelope. Both this pitch ambiguity and the phase randomization may have an impact on measures of TFS sensitivity and might impair subjects’ performance.

Here, we reexamine the TFS sensitivity of normal-hearing subjects using pulse spreading harmonic complexes (PSHCs; Hilkuysen & Macherey, 2014). PSHCs are harmonic complexes whose phase relationship is tailored to manipulate the envelope rate independently from the f_0 . In Experiment 1, we show that it is possible to produce two PSHCs that have the same long-term

spectrum, and identical envelopes but are perceived as either rising or falling in pitch across their duration, thereby avoiding the pitch ambiguity of inharmonic complexes. Acoustically, these stimuli only differ in their TFS. In Experiment 2, we use these stimuli to measure the lower limit of TFS sensitivity in different frequency regions and discuss the putative mechanisms underlying this limit.

Experiment 1: Pitch contour identification

Methods

Subjects and Equipment

Six normal-hearing subjects with ages ranging from 26 to 57 years (mean of 38.7) took part. Subject S4 was the author. They were seated in a double-walled sound proof booth in front of a computer screen. Stimuli were presented over an Etymotic ER2 earphone positioned inside one of their ears. The earphone was connected to a RME BabyFace Pro sound card. The experiment took place in the presence of a background noise that was played from a Zoom H4N recorder connected to the sound card.

Generation of PSHCs

The generation of PSHCs involves dividing the harmonics of a complex into several groups (called sub-complexes) and summing them separately before adding them back together with particular delays so that the overall envelope repetition rate remains regular across the stimulus duration. Following Hilkuysen & Macherey (2014), a PSHC is defined by

$$s(t/k) = \sum_{i=M}^N \sin(2\pi f_0 i t + \varphi(i, k)) \quad (1)$$

where f_0 is the fundamental frequency, M is the lowest harmonic present, i is the harmonic number, N is the highest harmonic present, and t is the time. The starting phase of harmonic i is given by

$$\varphi(i, k) = 2\pi \frac{i}{k^2} r_j \quad (2)$$

where k is the PSHC order which determines the envelope repetition rate (equal to $k^2 f_0$) and j is defined as

$$j = (i \bmod k) + 1 \quad (3)$$

The mod operator indicates the remainder after division of i by k . Equation (3) assigns the i^{th} harmonic to the j^{th} sub-complex so that there are in total k sub-complexes. In our previous publications (Hilkuysen & Macherey, 2014; Mesnildrey et al., 2015), r_j was a random element without replacement from $\{1, 2, \dots, k\}$ whose role was to spread the sub-complexes evenly across the stimulus period, presenting them in random order. Here, we consider two special cases of PSHCs where the sub-complexes are presented either in descending order (down-PSHC) or in ascending order (up-PSHC). For down-PSHC, r_j is given by

$$r_j = j \quad (4)$$

while for up-PSHC, it is given by

$$r_j = k - j + 1 \quad (5)$$

Stimuli

Stimuli consisted of up-PSHCs, down-PSHCs and inharmonic (INHARM) complexes synthesized in Matlab with a sampling rate of 96 kHz. Three versions of each of these stimuli were generated with envelope rates equal to 98 pulses per second (pps), 200 pps and 450 pps.

For the INHARM stimuli, the frequency components were randomly drawn with the constraint that the frequency separation between consecutive components was constant and equal to the envelope rate for each condition. For up- and down-PSHCs, the fundamental frequency was 2 Hz and the PSHC order k was 7, 10 and 15 for the three envelope rates, respectively.

All stimuli were bandpass filtered using a zero-phase 6th-order Butterworth filter (36 dB/oct. attenuation slopes) with -3-dB cut-off frequencies of 2000 and 2540 Hz, yielding a 2-ERB_N wide passband (Glasberg & Moore, 1990). The stimuli had a total duration of 500 ms (corresponding to a full period of the PSHCs) and were shaped with 20-ms raised-cosine onset and offset ramps. The PSHCs display gliding spectral ripples, as shown in Figure 1. For each presentation of the up- and down-PSHCs, the ripple starting phase varied randomly so that subjects could not identify the stimulus solely based on its onset or offset. The stimuli had an overall level 40 dB higher than the threshold in quiet of a pure tone with frequency corresponding to the middle of the passband. A continuous background of threshold-equalising noise (TEN) was presented at a level of 25 dB/ERB_N (same SNR as Moore & Sek (2009)).

The time waveforms and spectrograms of the three stimuli are shown in Figure 1 for an envelope rate of 450 pps. The blue curves in the top panels represent the time waveforms while the red curves show the Hilbert envelopes. At a given rate, the envelopes of the down- and up-PSHCs are identical but are slightly narrower and have a larger peak amplitude than the envelope of the INHARM complex (when compared at equal RMS). The spectrograms in the bottom panels show 2-seconds of the three stimuli. The down- and up-PSHCs exhibit linear spectro-temporal ripples. The ripple density is determined by the envelope rate which corresponds to the frequency separation between adjacent ripples, while the ripple velocity is determined by the f_0 . Given that the PSHCs contain only harmonics of 2 Hz, they are both periodic, exactly repeating every 500 ms. The shallow-slope filters applied to the stimuli make them similar to continuous Shepard tones (referred to as “Risset glissando”) except that in our case the frequency separation between adjacent ripples is equal to a constant value in Hz whereas the components in Shepard tones are uniformly spaced on an octave scale (Shepard, 1964; Risset, 1969).

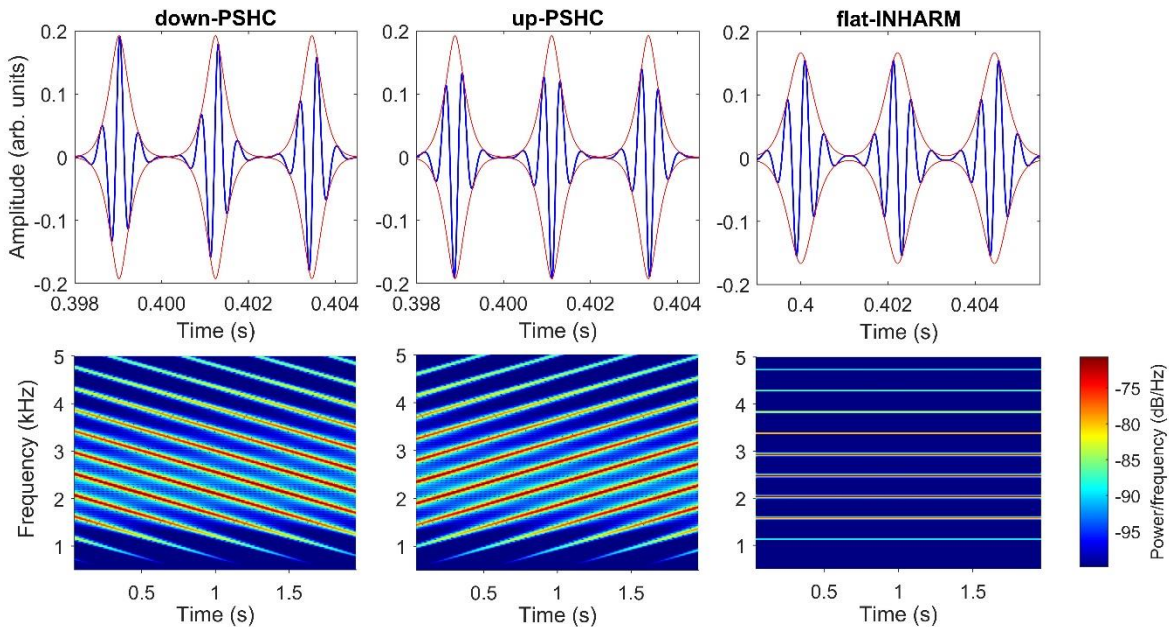


Figure 1 — Top: In blue: temporal waveforms of three stimuli used in Experiment 1, all at a rate of 450 pps. The red line shows the Hilbert envelope. Only 3 envelope periods are shown. Bottom: Spectrograms showing 2 seconds of the same three signals.

Procedure

The whole experiment was completed in one session lasting roughly an hour.

To determine the appropriate level of the complexes, the detection threshold was measured for a 2254-Hz (geometric center of the passband), 500-ms pure tone using a three interval, three alternative forced choice adaptive procedure with a 2-down, 1-up rule (c.f. Moore and Ernst, 2012). The signal level was adapted in steps of 8 dB until the second reversal, then in steps of 4 dB until the fourth reversal and finally in steps of 2 dB until the tenth reversal when the procedure stopped. The threshold was calculated as the mean signal level at the last six reversals. Two threshold estimates per subject were collected and their threshold was assumed to be the mean of these two estimates. As previously mentioned, the overall level of the complexes and the level of the background TEN used in the main part of the experiment were set at 40 dB and 25 dB/ERB_N above this pure tone threshold, respectively.

Experiment 1a used an identification task. On each trial, subjects heard one of the 9 stimuli (3 stimulus types x 3 envelope rates) and had to indicate whether they perceived the pitch as rising, falling or staying constant across the 500-ms stimulus. Feedback was not provided. All stimuli and envelope rates were presented in random order in blocks of 90 trials. Each subject performed a total of 5 blocks, yielding 50 trials per stimulus in total.

In Experiment 1b, the same subjects took part in a three-interval two-alternative, “odd-man-out” task to investigate whether they could discriminate between the up-PSHC, down-PSHC and INHARM stimuli at the lowest envelope rate (98 pps). They were presented with a reference stimulus in interval 1 followed by the same reference and a target stimulus randomly presented in intervals 2 and 3. There were three different reference-target trials (up-PSHC vs. down-PSHC, up-PSHC vs. INHARM and down-PSHC vs. INHARM). The trials were presented in blocks of 105, consisting of 75 trials at 98 pps (25 trials for each condition) and 30 easier “catch” trials (5 trials per condition for the two higher envelope rates). All subjects performed two blocks in total. Feedback was provided.

Results

The top panel in Figure 2 shows the percent correct scores obtained in Expt. 1a for the six subjects and the three envelope rates. Five out of six and six out of six subjects had scores higher than 92% for envelope rates of 200 and 450 pps, respectively, showing that their perception was, at these rates, consistent with expectations based on the time-frequency representations of the stimuli (Figure 1). For the lowest rate (98 pps), however, all subjects performed at chance. Further inspection of the data revealed that they perceived the stimuli to remain constant in pitch in 90% of trials (i.e., for the three stimulus types), indicating that both the down- and up-PSHCs were also perceived as constant in pitch.

The results of Expt. 1b (Figure 2, bottom panel) show that only S5 performed significantly above chance for discriminating the inharmonic complex from each of the two PSHCs (up vs. flat and down vs. flat conditions). However, none of the subjects could discriminate between the up- and down-PSHCs. This shows that at 98 pps, subjects were not sensitive to TFS. It is possible that S5 used remaining envelope differences between the inharmonic tone and the PSHCs to perform the task. The observation that he could not discriminate between the two PSHCs which had identical envelopes would argue in favour of this explanation.

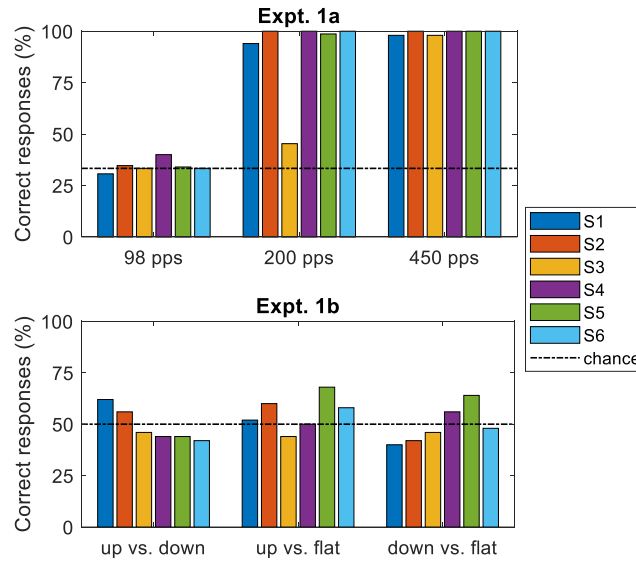


Figure 2 — Top: Results of Expt. 1a showing the percentage of correct identification at each of the three envelope rates for each subject. Bottom: Results of Expt. 1b showing the percentage of correct discrimination of the three pairs of stimuli for each subject.

Discussion

The results of Experiment 1 show that, for a given passband, there is an envelope rate below which down- and up-PSHCs cannot be discriminated and are not being perceived as changing in pitch. Below we describe the cues that the subjects may have used to perform the pitch contour identification task of Expt. 1a and discuss the possible reasons why their performance got poorer at the lowest envelope rate. These cues are excitation pattern cues and phase locking cues.

The top panels of Figure 3 show simulations of basilar membrane vibrations using a Gammatone filterbank (Slaney, 1998) in response to 2-seconds of 450-pps envelope rate, down- and up-PSHCs in the absence of the background TEN. Each line represents the evolution of the RMS value of a certain auditory filter output (whose characteristic frequency is indicated on the y axis) computed across successive periods of the envelope. There is a clear amplitude modulation at the output of all filters at a frequency equal to the fundamental frequency (2 Hz). The starting phase of this modulation is, however, different across filters so that at any given time, some filters have a high amplitude and others have a low amplitude. This is illustrated in the bottom panels of Figure 3 which show excitation patterns calculated across a single envelope period at three successive times across the stimulus. These excitation patterns represent the RMS outputs of the Gammatone filters expressed on a dB scale. All excitation patterns show peaks separated by a frequency equal to the envelope rate. However, the peaks drift during the stimulus period. For the down-PSHC, the peaks shift progressively downwards across the stimulus duration while for the up-PSHC, they shift upwards. The perception of down-PSHCs as falling in pitch and of up-PSHCs as rising in pitch may reflect the use of excitation pattern cues in two different ways: Subjects may either selectively attend a particular peak in the excitation pattern and follow its course across the stimulus duration or they may perceive a continuously changing residue pitch based on the continuous analysis of the pattern of peaks present in the excitation pattern (e.g., using a pattern-matching mechanism based on central templates; c.f. Terhardt, 1974). When the envelope rate is decreased, this has the effect of increasing the ripple density and, therefore, reducing the size of the peaks in the excitation pattern. This is similar to what happens when decreasing the f_0 of a harmonic complex filtered in a fixed frequency region: the peaks in the excitation pattern get smaller and smaller as the harmonics become less and less resolved. This may, therefore, explain why the identification task was impossible at 98 pps.

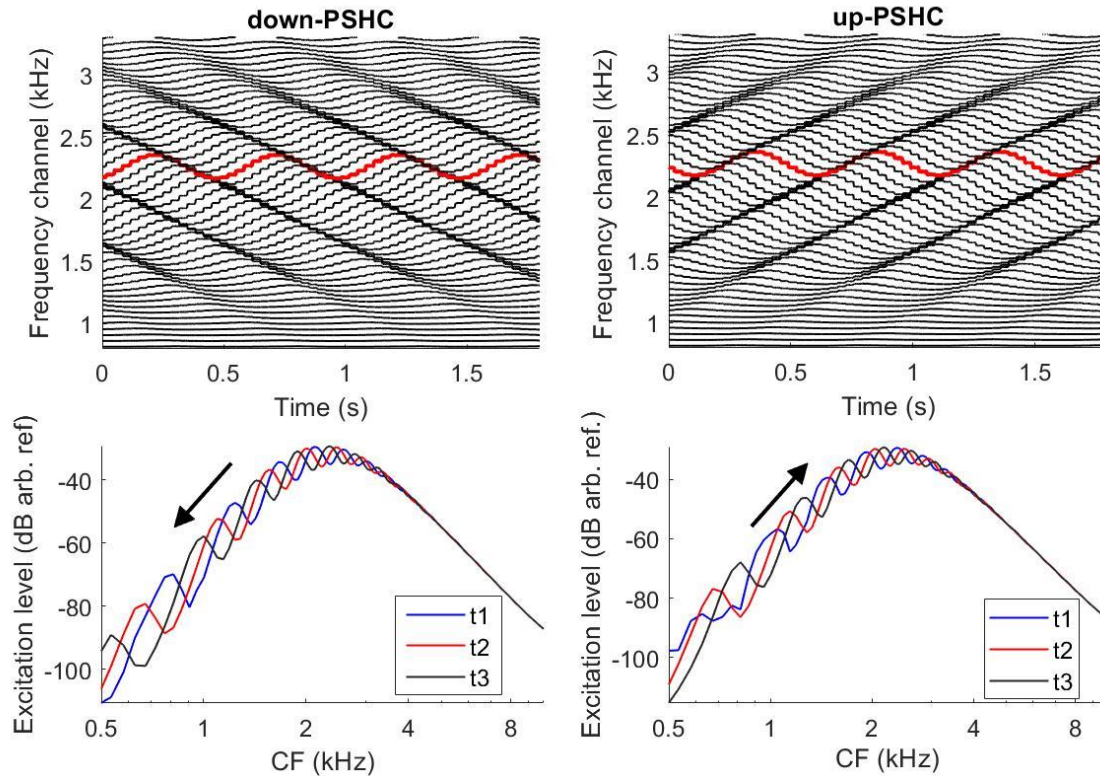


Figure 3 — Top: RMS energy at the output of several Gammatone filters in response to down- and up-PSHCs with an envelope rate of 450 pps. The RMS energy was calculated across successive envelope periods. The red function corresponds to the filter located in the center of the passband (2254 Hz). Bottom: Excitation patterns at three successive times (t1, t2 and t3) during a period of the stimulus for the same signals.

The second potential mechanism is temporal. The top panels of Figure 4 show the output of an auditory filter centred on 2254 Hz over 2 seconds (4 periods of the f_0) for both PSHCs. The similarity between the outputs obtained for the two PSHCs suggest that there is no information in the envelope at the output of a given auditory filter that can allow the identification of the pitch contour. The middle panels of Figure 4 show a few envelope periods of these same signals. It can be observed that the TFS changes from one envelope period to the next. To quantify this change and to determine the time intervals conveyed by the TFS peaks, the auditory filter output waveform of the first envelope period of the stimulus was cross-correlated with the output waveform of the second envelope period. The lags corresponding to the first three maxima of the cross-correlation function were stored. Then, the second period was cross-correlated with the third period and so on until the end of the stimulus. The bottom panels of Figure 4 show the reciprocal of the lags corresponding to the three highest peaks of these cross-correlation functions as a function of time. These values can be viewed as the possible “pitches” conveyed by the TFS peaks at the output of this specific filter at any given time. This analysis shows that the TFS information *within* a single auditory filter provides a cue that may allow subjects to perceive the pitch contour of down- and up-PSHCs because the reciprocal of the distance between the TFS peaks progressively decreases for down-PSHC and progressively increases for up-PSHC. However, these TFS variations are still present at low envelope rates (not shown here), including at 98 pps where subjects could not do the task. Therefore, to explain why performance is getting worse at low rates, we also need to assume that there is a frequency-dependent temporal limitation on the time intervals that can be analyzed by the central auditory system to extract pitch, as already proposed in previous publications (Moore, 1982; Pressnitzer et al., 2001).

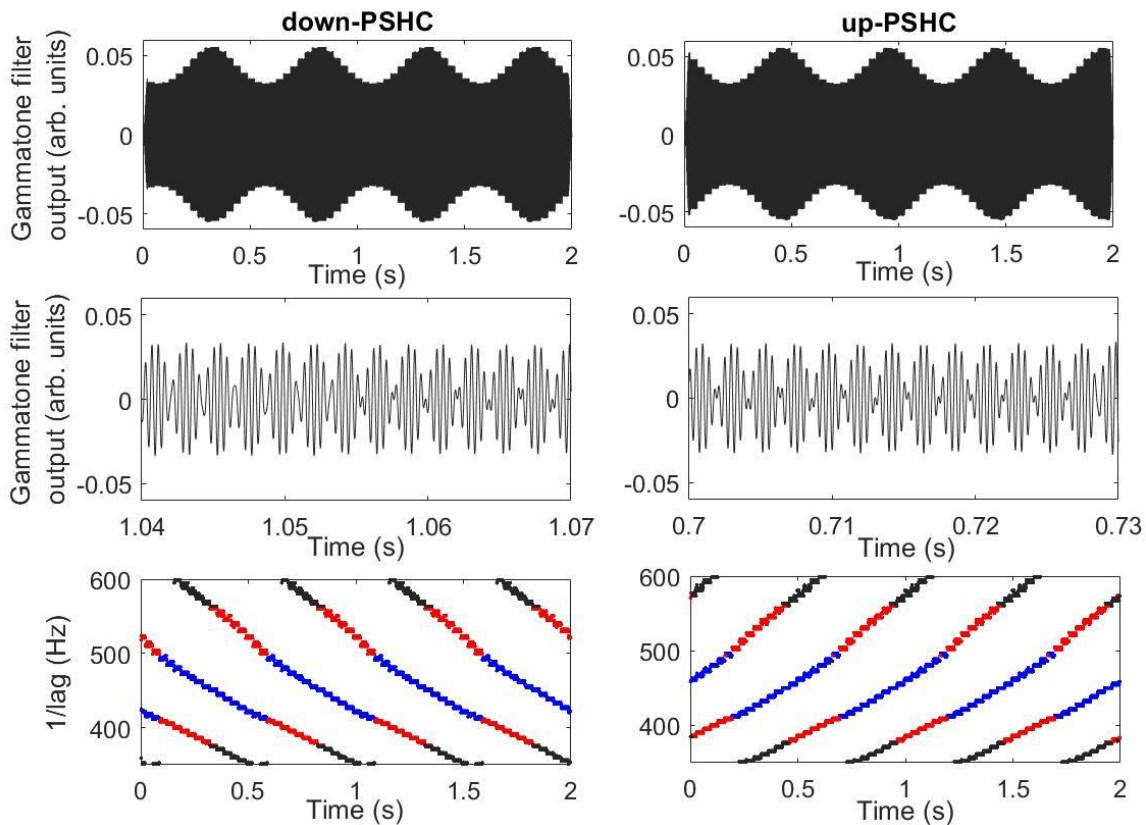


Figure 4 — Top panels: Output of a gammatone filter centered on 2254 Hz (center of the passband) for the two PSHCs. Middle panels: Short portions of the same output showing the TFS. Bottom panels, Reciprocal of the lags corresponding to the three highest peaks of the cross-correlation function between one envelope period and the next. Blue, red and black symbols are for the highest, second-highest and third-highest peaks, respectively.

Experiment 2: Measuring the lower limit of TFS sensitivity

Methods

Subjects

Ten subjects with ages ranging from 21 to 41 (mean of 25.5) took part in Experiment 2, including one subject (the author) who had participated in Experiment 1. All subjects were musicians and/or Master's students in Acoustics. They were selected after verifying they had normal absolute hearing thresholds up to 14 kHz. Pure tone detection thresholds at frequencies of 0.25, 0.5, 1, 2, 4, 8, 11.2 and 14 kHz were measured using the same adaptive task as for Expt. 1.

Stimuli

The stimuli were down- and up-PSHCs of various envelope rates filtered into seven different frequency regions. The filtering was identical to that for Expt. 1 except that F_c was equal to 0.25, 0.5, 1, 2, 4, 8 or 11.2 kHz depending on condition. The upper cut-off frequency was set so that the passband width would always be equal to 2 ERB_N . The overall presentation level of the complexes was again 40 dB higher than the pure tone detection threshold for a frequency corresponding to the centre of the passband and the level of the TEN noise was 25 dB/ ERB_N above the detection threshold. Because we did not measure the thresholds in the middle of each passband, this reference threshold was obtained by interpolation from the threshold measurements made for the different values of F_c .

Procedure

The main task was a 3I2AFC, 2-down, 1-up adaptive task. In the first interval, the subjects always heard the down-PSHC. In intervals 2 and 3, the up-PSHC and down-PSHC were presented in random order. Subjects had to indicate which stimulus was different than the other two. The three stimuli presented in each trial had the same envelope rate, so subjects could only use differences in TFS to do the task. The starting envelope rate was equal to $0.2F_c$. Following two correct responses, the envelope rate was divided by $(1.05)^\alpha$ while following one incorrect response, it was multiplied by $(1.05)^\alpha$. α was equal to 4 until the second reversal, then it was switched to 2 until the fourth reversal and finally to 1 until the tenth reversal at which point the procedure stopped.

While we aimed to use the same f_0 of 2 Hz for all stimuli, this was not always possible because, for a given f_0 , the envelope rate can only be equal to $k^2 f_0$ where k is an integer and can therefore only take a finite number of values. Therefore, the f_0 value closest to 2 Hz was selected for each desired envelope rate. This means that the f_0 could slightly vary from trial to trial (from 1.7 to 2.4 Hz in the lowest frequency region and substantially less in higher frequency regions). Note, however, that the f_0 was constant *within* a trial and could not be used as a cue to perform the task.

A measurement block involved collection of seven discrimination threshold measurements, one for each frequency region, presented in random order. Subjects performed between 5 and 6 blocks in total and the first block was considered as training and was not taken into account in the analyses. This experiment was completed in two sessions of 2 hours.

Results

Figure 5A shows the discrimination thresholds (i.e. reflecting the lower limit of TFS sensitivity) obtained as a function of F_c . The lower limit of TFS significantly increased with increases in F_c , ranging from 42 pps for $F_c=0.25$ kHz to 691 pps for $F_c=11.2$ kHz. Except for one subject whose threshold was lower at 0.5 than at 0.25 kHz, all subjects showed a monotonic increase as a function of F_c .

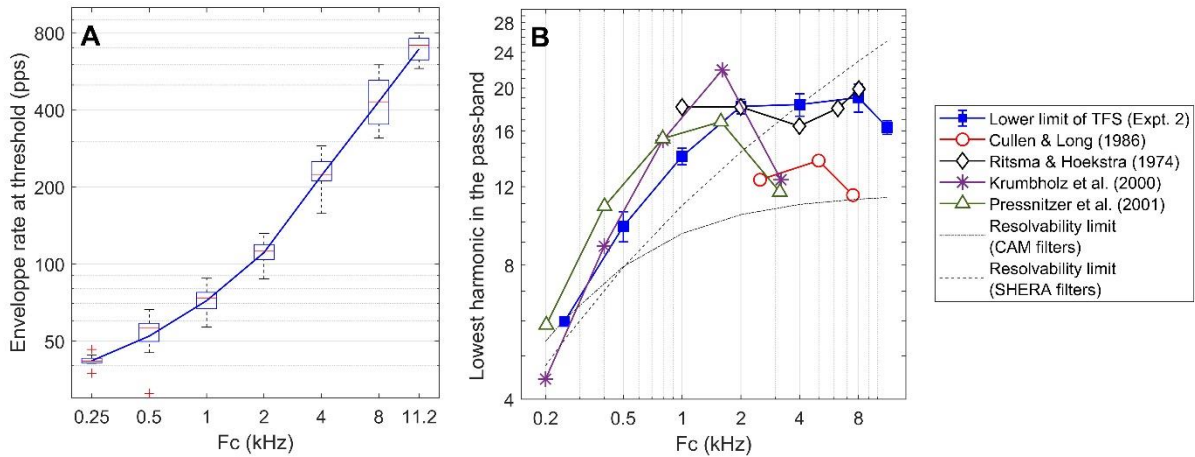


Figure 5 — A: Boxplots showing the distributions of the lower limit of TFS sensitivity for each F_c . Each boxplot shows the median, 25th and 75th percentiles and extreme values. Outliers (red '+' symbols) correspond to values that are more than 1.5 times the interquartile range away from the bottom or top of the box. The solid line shows the mean across subjects. B, The same data as in panel A are plotted in blue as a function of the lowest virtual harmonic in the passband (see main text). The dashed and dashed-dotted lines show the limit of harmonic resolvability for two sets of auditory filters (see main text for the definition of resolvability). The other functions illustrate transition rates from good to poor frequency discrimination for harmonic complex tones, replotted from Krumbholz et al. (2000).

The solid blue line connecting the filled squares in Figure 5B shows the same data expressed as the ratio between F_c and the lower limit of TFS. We refer to this dimensionless variable as the rank of the lowest *virtual* harmonic present in the passband “ N_{LOW} ” at threshold as it represents the rank of the lowest-frequency ripple present in the passband. A one-way ANOVA showed a significant effect of F_c on N_{LOW} ($F(6,54)=43$, $p<0.001$). Paired t-tests (without correction for multiple comparisons) performed on the data obtained for successive values of F_c showed that N_{LOW} increased significantly from 0.25 to 0.5 kHz ($p=0.001$), from 0.5 kHz to 1 kHz ($p=0.003$) and from 1 kHz to 2 kHz ($p<0.001$), remained approximately constant from 2 to 8 kHz and finally decreased from 8 to 11.2 kHz ($p=0.042$). Interestingly, the across-subject mean N_{LOW} value was higher than 18 for $F_c=2, 4$ and 8 kHz and was still higher than 16 for $F_c=11.2$ kHz. These large N_{LOW} values suggest that the ripples were spectrally unresolved in the passband according to usual definitions of resolvability (Moore & Gockel, 2011).

Discussion

As previously explained, our subjects may have used either excitation pattern or phase locking cues to perform the tasks in the two experiments. If subjects based their judgments on excitation pattern cues, we would expect the lower limit of the use of TFS to depend on the bandwidths of the auditory filters. The dashed line of Figure 5B illustrates the limit of harmonic resolvability as defined by Moore & Ohgushi (1993); components separated by less than $1.25ERB_N$ are assumed to be unresolved. This limit falls below the lower limit of TFS sensitivity for all regions except the lowest, suggesting that, at threshold, the ripples of our stimuli were unresolved. The dash-dotted line also shows the same “resolvability” limit, but using ERB values proposed by Shera et al. (2002), which are smaller at high frequencies. These ERB values likely represent a conservative definition of resolvability given that they were obtained at lower levels than used here (Oxenham & Simonson, 2006). For $F_c<4$ kHz, the lower limit of TFS still falls above this line, confirming that the ripples at threshold were unresolved even assuming narrow auditory-filter bandwidths. For $F_c>4$ kHz, it cannot be excluded that the ripples were partially resolved. However, if the subjects based their judgments on excitation pattern cues, we would expect N_{LOW} for the lower limit of TFS sensitivity to continue to increase up to the highest F_c . This was not the case as N_{LOW} remained mostly constant above 2 kHz. Also, it should be noted that the random fluctuations produced by the background TEN should reduce the ability to use excitation-pattern cues (Jackson & Moore, 2014). Thus, unless our subjects managed to use residual excitation pattern cues in the lower skirt of the filter, these data suggest that the lower limit of the use of TFS is related to the use of temporal (phase-locking) cues rather than excitation-pattern cues.

Moore et al. (2009) measured sensitivity to TFS for several f_0 values in different frequency regions. They showed that for stimuli bandpass filtered so that the lowest component in the passband, N_{LOW} , was harmonic 13, subjects could discriminate harmonic and inharmonic tones for $f_0=50, 100$ and 200 Hz. At these f_0 , harmonic 13 corresponds to $F_c=1.3, 2.6$ and 5.2 kHz, respectively. Our results are consistent with their data because at these F_c values, our subjects could always discriminate between down- and up-PSHCs when N_{LOW} was the 14th or lower (see Figure 5B). They also showed that subjects required the presence of lower harmonics to perform the same task for $f_0=35$ Hz. Similarly, Figure 5B shows that our N_{LOW} was lower in low frequency regions. Tarnowska et al. (2019) showed above-chance performance for musicians on a variant of the TFS task used in Moore et al. (2009) for $F_c=3.8$ kHz and $N_{LOW}=19$ which is also consistent with the present data. Finally, Moore & Sek (2009) extended these measurements to higher fundamental frequencies ($f_0=800$ and 1000 Hz) and showed sensitivity to TFS when N_{LOW} was equal to 12 although only 6 out of their 8 subjects could perform the task at the highest f_0 . Their stimuli are equivalent to $F_c=9.6$ and 12 kHz, respectively, and can be compared to our highest-frequency region ($F_c=11.2$ kHz) for which subjects could discriminate down- and up-PSHCs for $N_{LOW}=16$ and lower. It is possible that the poorer performance Moore & Sek (2009) observed for $F_c=12$ kHz than for $F_c=9.6$ kHz is related to the degradation of phase locking cues at very high frequencies. The trend of decreased performance we observed in the highest frequency region may have the same origin.

In summary, the finding that subjects were still sensitive to TFS when N_{LOW} was equal to 18 for $F_c=2, 4$ and 8 kHz suggests that the ripples were unresolved and that subjects used temporal cues to perform the task. These results, therefore, confirm the findings of Moore & Sek (2009) and add further evidence that phase locking cues can be used by humans at very high frequencies.

Finally, it has been shown that the frequency discrimination of harmonic complex tones depends on the rank of the lowest harmonic present in the passband. When the frequency region is fixed, f_0 difference limens (DLs) are poorer at low than at high f_0 (Ritsma & Hoekstra, 1974; Cullen & Long, 1986; Krumbholz et al., 2000). The transition from good to poor performance, however, depends on frequency region. Krumbholz et al. (2000) compiled the results of several of these studies (c.f. their Figures 7 and 10) and defined the transition rate as the rate at which the f_0 DL equals 2.5% except for the Ritsma and Hoekstra (1974) data for which the criterion was set to 1%. These transition rates, expressed in terms of the rank of the lowest harmonic number in the passband are replotted in Figure 5B. The upward triangles also show the lower limit of melodic pitch defined as the rate at which subjects could no longer recognize that the f_0 of a complex tone in a melody changed by a semitone (Pressnitzer et al., 2001). The functions showing these transition rates as a function of F_c follow a similar pattern as the lower limit of TFS, first showing an increase followed by a decrease or a flat portion above 1 kHz. Although speculative, it is possible that the transition from good to poor f_0 discrimination of complex tones with increasing harmonic rank reflects the decreased sensitivity to TFS in high-frequency regions.

Conclusions

These experiments show that it is possible to generate pairs of harmonic complexes that share the same long-term spectrum and the same temporal envelope but have different TFS and elicit different pitch percepts. These down- and up-PSHCs add to the arsenal of tools for studying sensitivity to TFS. The lower limit of TFS sensitivity (expressed as the envelope rate below which subjects were insensitive to TFS) was shown to increase with increasing frequency region. The observation that subjects were still sensitive to TFS in high frequency regions (> 4 kHz) when the stimuli likely contained only unresolved components suggests that phase locking cues remain usable at high frequencies, confirming the results of previous studies.

Acknowledgements

I thank Bob Carlyon and Brian Moore for their comments on a previous version of this manuscript.

References

- Cullen, J. K., Jr, & Long, G. R. (1986). Rate discrimination of high-pass-filtered pulse trains. *The Journal of the Acoustical Society of America*, 79(1), 114–119. <https://doi.org/10.1121/1.393762>
- Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1-2), 103–138. [https://doi.org/10.1016/0378-5955\(90\)90170-t](https://doi.org/10.1016/0378-5955(90)90170-t)
- Hilkhuyzen, G., & Macherey, O. (2014). Optimizing pulse-spreading harmonic complexes to minimize intrinsic modulations after auditory filtering. *The Journal of the Acoustical Society of America*, 136(3), 1281. <https://doi.org/10.1121/1.4890642>
- Jackson, H. M., & Moore, B. C. (2014). The role of excitation-pattern, temporal-fine-structure, and envelope cues in the discrimination of complex tones. *The Journal of the Acoustical Society of America*, 135(3), 1356–1370. <https://doi.org/10.1121/1.4864306>

- Köpl C. (1997). Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, *Tyto alba*. *The Journal of neuroscience*, 17(9), 3312–3321. <https://doi.org/10.1523/JNEUROSCI.17-09-03312.1997>
- Krumbholz, K., Patterson, R. D., & Pressnitzer, D. (2000). The lower limit of pitch as determined by rate discrimination. *The Journal of the Acoustical Society of America*, 108(3 Pt 1), 1170–1180. <https://doi.org/10.1121/1.1287843>
- Mesnildrey, Q., Hilkuysen, G., & Macherey, O. (2016). Pulse-spreading harmonic complex as an alternative carrier for vocoder simulations of cochlear implants. *The Journal of the Acoustical Society of America*, 139(2), 986–991. <https://doi.org/10.1121/1.4941451>
- Moore, B. C. J. (1982). *An Introduction to the Psychology of Hearing*, 2nd Ed. (Academic Press, London), pp. 293.
- Moore, B. C., & Ohgushi, K. (1993). Audibility of partials in inharmonic complex tones. *The Journal of the Acoustical Society of America*, 93(1), 452–461. <https://doi.org/10.1121/1.405625>
- Moore, B. C., & Sek, A. (2009). Sensitivity of the human auditory system to temporal fine structure at high frequencies. *The Journal of the Acoustical Society of America*, 125(5), 3186–3193. <https://doi.org/10.1121/1.3106525>
- Moore, B. C., & Gockel, H. E. (2011). Resolvability of components in complex tones and implications for theories of pitch perception. *Hearing research*, 276(1-2), 88–97. <https://doi.org/10.1016/j.heares.2011.01.003>
- Moore, B. C., & Ernst, S. M. (2012). Frequency difference limens at high frequencies: evidence for a transition from a temporal to a place code. *The Journal of the Acoustical Society of America*, 132(3), 1542–1547. <https://doi.org/10.1121/1.4739444>
- Moore, B. C., Hopkins, K., & Cuthbertson, S. (2009). Discrimination of complex tones with unresolved components using temporal fine structure information. *The Journal of the Acoustical Society of America*, 125(5), 3214–3222. <https://doi.org/10.1121/1.3106135>
- Oxenham, A. J., & Simonson, A. M. (2006). Level dependence of auditory filters in nonsimultaneous masking as a function of frequency. *The Journal of the Acoustical Society of America*, 119(1), 444–453. <https://doi.org/10.1121/1.2141359>
- Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A., & Santurette, S. (2011). Pitch perception beyond the traditional existence region of pitch. *Proceedings of the National Academy of Sciences of the United States of America*, 108(18), 7629–7634. <https://doi.org/10.1073/pnas.1015291108>
- Pressnitzer, D., Patterson, R. D., & Krumbholz, K. (2001). The lower limit of melodic pitch. *The Journal of the Acoustical Society of America*, 109(5 Pt 1), 2074–2084. <https://doi.org/10.1121/1.1359797>
- Risset, J.-C. (1969). Pitch control and pitch paradoxes demonstrated with computer-synthesized sounds. *J. Acoust. Soc. Am.* 46, 88. <https://doi.org/10.1121/1.1973626>
- Ritsma, R. J., and Hoekstra, A. (1974). "Frequency selectivity and the tonal residue," in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer, Berlin), pp. 156–163.
- Shepard, R. N. (1964). Circularity in judgments of relative pitch. *J. Acoust. Soc. Am.* 36, 2346–2353. <https://doi.org/10.1121/1.1919362>
- Shera, C. A., Guinan, J. J., Jr, & Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proceedings of the National Academy of Sciences of the United States of America*, 99(5), 3318–3323. <https://doi.org/10.1073/pnas.032675099>
- Slaney, M. (1998). Auditory toolbox. Interval Research Corporation, Tech. Rep 10, 1194

Tarnowska, E., Wicher, A., & Moore, B. (2019). The effect of musicianship, contralateral noise, and ear of presentation on the detection of changes in temporal fine structure. *The Journal of the Acoustical Society of America*, 146(1), 1. <https://doi.org/10.1121/1.5114820>

Terhardt E. (1974). Pitch, consonance, and harmony. *The Journal of the Acoustical Society of America*, 55(5), 1061–1069. <https://doi.org/10.1121/1.1914648>

Verschooten, E., Shamma, S., Oxenham, A. J., Moore, B., Joris, P. X., Heinz, M. G., & Plack, C. J. (2019). The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. *Hearing research*, 377, 109–121. <https://doi.org/10.1016/j.heares.2019.03.011>

Zwislocki, J., & Feldman, R. S. (1956). Just Noticeable differences in dichotic phase. *The Journal of the Acoustical Society of America*, 28(5), 860–865. <https://doi.org/10.1121/1.1908495>