

MISSING FUNDAMENTALS: A PROBLEM OF AUDITORY OR MENTAL PROCESSING?

Henning Reetz

Allgemeine Sprachwissenschaft
University of Konstanz
D-78464 Konstanz, Germany

Phone: +49 7531 882928, FAX: +49 7531 883095, E-mail: henning.reetz@uni-konstanz.de

ABSTRACT

Subjects were presented with signal pairs with different musical intervals. Signals were sine tones, complex tones with a fundamental, and complex tones without a fundamental. Subjects had to decide, which signal pairs form a specific musical interval. Reaction times indicate that the perception of the ‘missing fundamental’ is a sort of musical processing and not necessarily a part of normal auditory processing in pitch perception.

1. BACKGROUND

Pitch perception is always part of speech perception. Research on speech perception normally tries to single out individual aspects, like the segmental structure of a speech signal, its intonation, aspects of the duration, spectrum, etc. For the investigation of perceived pitch, psychoacoustic work is usually performed with more or less artificial signals to tease apart individual ingredients which give rise to a certain pitch percept. The underlying assumption, though normally not made explicit, is that the pitch processing in the brain is performed on a rather ‘low’ level even if it happens in some central ‘pitch processor’ ([1], [2], [3]) and can be described as auditory processing, independent of any ‘higher’ knowledge about speech, language, etc. The perception of intonation in this framework then builds on this auditory pitch perception.

It is assumed in psychoacoustic research that experiments in pitch perception measure the auditory performance of a listener. The processes in a person’s brain involving the perception of the pitch of a speech signal (e.g. as the intonation contour of a phrase) are presumably based on this auditory pitch perception mechanism. A well known effect is the perception of the ‘missing fundamental’, which gives an insight into fundamental processes of pitch perception.

One observation often made in psychoacoustic experiments on pitch perception is that only a few

subjects are able to perform the task (e.g. [1], [4], [5], [6]), and that listeners often make ‘misperceptions’ (e.g. [7], [8]). This is a surprising observation considering that there are no tone deaf listeners and that all listeners can follow an intonation contour with ease in a normal conversation. Therefore, the question arises of whether the psychoacoustic tasks measure the auditory pitch perception at all, or whether they measure a sort of ‘musical performance’ or some other higher sort of mental processing which is not part of the normal auditory processing and probably not part of the pitch perception in speech. The experiments reported here try to answer the question “how relevant are psychoacoustic pitch perception experiments for speech perception models?”

In this study the reaction times in a decision task are investigated in relation to the stimuli. It is assumed that decisions made in early stages of processing (auditory pitch perception) lead to shorter reaction times while decisions based on higher levels of mental processing (musical knowledge, intellectual decisions, etc.) lead to longer reaction times. The length of reaction is used as an indicator for which level of the cerebral processing (early = auditory, later = musical/intellectual) decisions in psychoacoustic experiments take place.

One of the most basic experiments in psychoacoustic pitch perception is the perception of the missing fundamental. The claim is that the fundamental frequency of a complex signal is perceived as pitch of the signal even if this frequency is removed from the spectrum ([9], [10]). There are a number of follow-up experiments which investigate the nature of this perception in more detail, and they all explain that the perceived pitch is either an effect of the periodicity pattern of the signal or is computed in some way from its harmonics (see [11], [12] for an overview). The underlying assumption is always that the perception of the missing fundamental is a basic process in pitch perception. To investigate this seemingly obviously correct assumption, experiments described in §2 were conducted.

2. EXPERIMENTS

The assumption behind this experiment is that the auditory perception of pitch comes prior to any musical or speech related processing. In other words, it is assumed that the auditory processing of a signal delivers some pitch representation, which is then further processed and which is perceived as, e.g., musical notes or intonation. Consequently, auditory perception of pitch has to be faster than musical or speech perception of pitch. If a listener has to react to presented signals, a reaction based on auditory processing should be faster than a reaction based on some later processing. Precisely this assumption is the rationale behind the experiments.

2.1 Material

Pairs of signals from the musical scale between c1 (261.6 Hz) and h2 (493.9 Hz) were generated (48 kHz sampling rate). All signals were 300 ms long, had a 20 ms sigmoidal on- and off-ramp, and were matched in RMS-amplitude. The signals in each pair were either sine tones (S), complex tones with a fundamental (C, 12-tone complex with 6 dB roll-off), or complex tones without a fundamental, a 'Missing Fundamental' (M, 11-tone complex with 6 dB roll-off). This results in 9 pair-combinations (CC, MM, SS, CS, SC, CM, MC, MS, SM). The signals pairs were either one octave apart, a third apart (i.e. 4 half-tones), formed a prime (i.e. were equal in frequency), or formed another musical interval not further apart than one octave. Half of the intervals (not including the primes) were rising and the other half were falling. From all possible pair-combinations 50 pairs each of primes, upward Octaves, downward octaves, upward thirds, and downward thirds were selected, and added to 125 pairs each with other upward and downward intervals. The resulting 500 signal pairs were randomized and copied onto a digital audio tape (DAT) in the following manner. Each trial was made up of a faint attention click, a 300 ms pause, the first signal of a pair, a 250 ms pause, the second signal, and a pause of 3500 ms in which the subjects were asked to react. Reaction time measurement started with the presentation of the second signal in a pair. The total length of the experimental tape was 38 minutes.

2.2 Subjects and task

5 subjects with professional musical training who were musical performers on a semi-professional level served as subjects of the experiment. The subjects' task was to press the right button of a two-button panel in front of them if they hear a difference of an octave (12 half-tones), a third (4 half-tones), or a prime (identical frequency), and to press the left button if they hear another interval. For left-handed persons the instructions and the buttons were

reversed. Consequently, in the subsequent text all subjects are treated as being right-handed.

The reaction (button pressed and reaction time) was recorded by appropriate hardware in the linguistic lab of the Konstanz University. Reaction time measurement started at the beginning of the second tone of a pair.

The subjects were instructed about their task and a tape similar to the experimental tape was presented to them for several minutes as exercise. After that, the experimental tape was presented to them at appropriate hearing level via headphones (Sennheiser HD 520II) without interruption.

3. RESULTS

The subjects reported that they found the task interesting and not very complicated, but that they had the impression that they sometimes had to 'think' about their decisions. (In a prototype experiment with shorter time-out times to prevent subjects from thinking about the stimuli, the one subject tested often missed reactions; therefore, the longer pause after the second signal was chosen.) In particular, they found it more complicated to decide on the downward movements, especially for the thirds. Some of the subjects reported that they sometimes heard a 'note one octave higher than a base note, to which the third was played'. These cases can not be traced with the set-up of the experiment, but it can be assumed that these are cases of thirds in which one of the signals in the pairs had a missing fundamental.

Between 1% (octave and third, upwards) and 4% (third, downwards) errors occurred for non-identical intervals, but 16% of the signals with identical F0 led to errors. In the latter case, the subjects hardly ever made errors if the types of the signals (S, C, M) were identical, but errors were around 30% if C and M had to be compared and around 6% if S was one of the signals. Apparently, the two complex signals were more often not perceived as having the same F0, but did not lead to a problem if the interval had to be scaled as a musical interval (the prime is a musical interval, but it is primarily a decision on identity that has to be performed, rather than a musical task).

The difference in the reactions between same or different signal types for primes appear also in the average reaction times of the subjects (see Fig. 1). In this graph, the reaction times for correct reactions (i.e., right buttons for prime, third, and octave reactions, left buttons otherwise) are broken down for interval and direction, if appropriate. As to be expected, the reactions were fastest if two signals of the same type and frequency had to be compared (284 ms). As mentioned above, in this case errors were hardly ever produced. However, comparing signals of different type with identical frequency took nearly twice as long (526 ms), and longer than many of

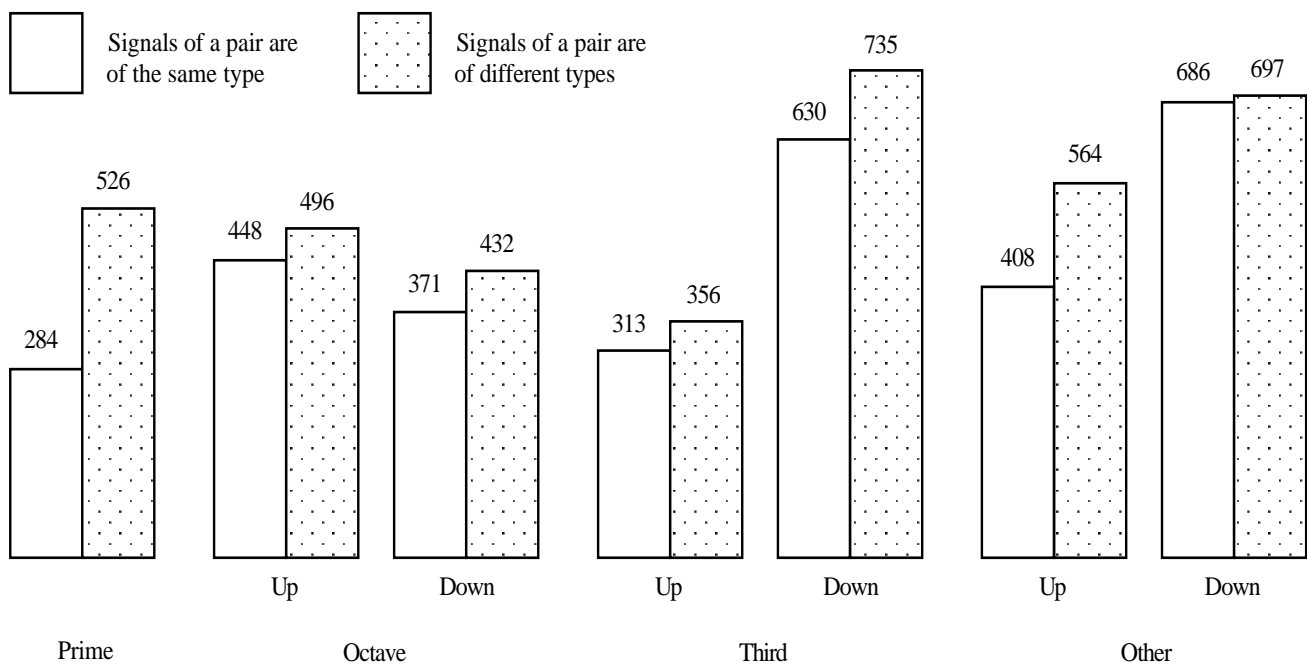


Figure 1: Average reaction times (in ms) to compare two signals that are a prime, octave, third, or other interval apart (up or down), broken down for pairs of the same type (CC, MM, SS) or of different type (CS, SC, CM, MC, MS, SM).

the other interval decisions. This large difference in reaction times between signal pairs of the same type and of different types does not show up for other intervals, although the comparison of different signal types took in general longer than the comparison of the same signal types. Comparing upward octaves took 448 and 496 ms, resp., while downward comparison had similar times of 371 and 432 ms, resp. Thirds, on the other hand, show a large difference between upward (313 and 356 ms) and downward (630 and 735 ms) intervals. A large difference can be observed between the reaction times for up- and downward movements of other intervals if both signals were of the same type (408 vs. 686 ms), but the difference is smaller if the two signals were of different types (564 vs. 697 ms). It must be kept in mind, that the 'other' decision had to be performed by the subjects with their less-preferred hand, i.e., the reaction times in this case are expected to be slower than for the preferred hand decisions.

From these results, at least three observations can be made: (1) frequency identity decisions depend strongly on the identity of the signal type; (2) third upward decisions are markedly faster than downward decisions; (3) in the 'other intervals' class, upwards decisions with identical signal types are much faster than upwards decisions with different signal types. As already mentioned, the first observation is not surprising considering that it is a simple 'identity' decision about two auditory events – no musical knowledge is involved. This case should serve as baseline for the other comparisons. Somewhat surprising are the reaction times for the thirds upward decisions, which are nearly as fast as the identity decisions. Asking some subjects after the experiment about this case, they

commented that they found an upward third a rather 'natural' interval, which often occurs in music and which, especially in the case of the upward interval, is used in professional ear-training. They also reported that they can easily identify this interval even with an octave interspersed, while the downward movement took some internal processing, especially if they heard one of the notes one octave higher. Regarding the third observation, nothing can be said at the moment about it and more investigation is required to sort out whether there are slower and faster conditions in this class. It is well possible that some (but not all) of the 'other' intervals are easily perceived (like the third upwards) and identified as the 'interval X'. If this is the case, the data must be analyzed for the individual musical intervals separately and cannot be treated as one group.

Breaking down the results in terms of the relation between the class of tones repeats the finding, as to be expected, that the comparison of two tones from the same class is fast (CC: 479 ms, MM: 491 ms, SS: 473 ms). Comparing a complex signal with a fundamental and a sine tone in the sequence CS is just as fast (482 ms), while the reversed sequence takes more time (522 ms). The differences in reaction times between the order of presentation is considerably larger if a complex without a fundamental is involved (CM: 554 ms vs. MC: 656 ms and MS: 526 ms vs. SM: 682 ms). The finding that the presentation of a complex signal before a simpler signal seems to be easier for a subject is in accordance with earlier findings [8], where musically 'naive' subjects had to compare speech signals with sine tones. The reasons for this behaviour can be different in the comparisons where the

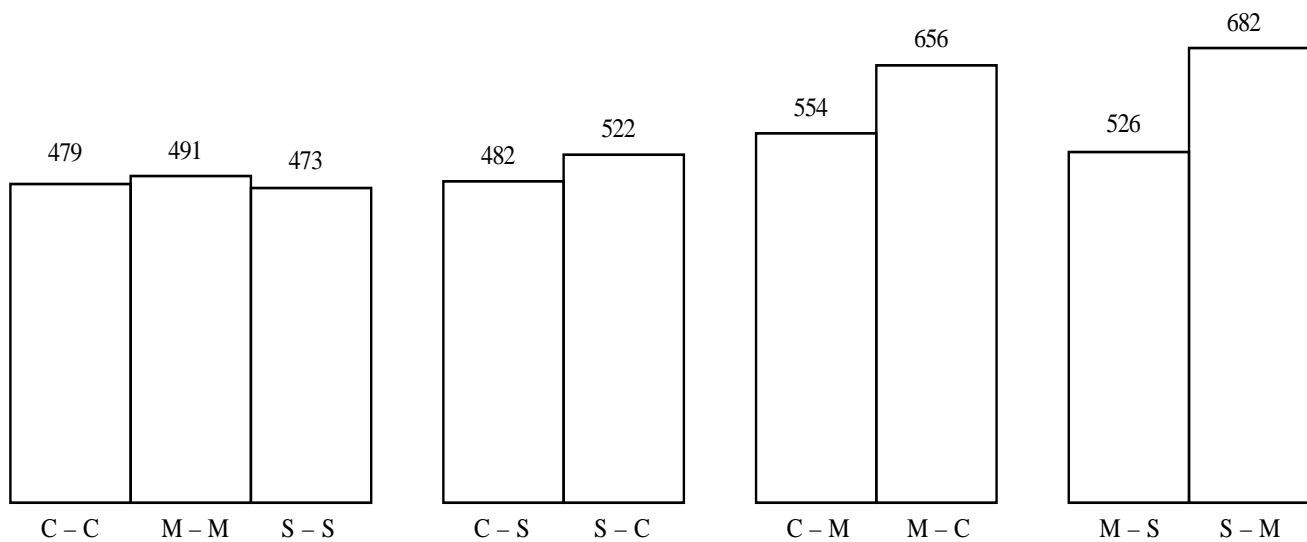


Figure 2: Reaction times (in ms) to compare signals of the indicated type in a pair (S: sine tones, C: Complex signals with a fundamental, M: Complex signals without a fundamental).

complex signals and where the sine tones are involved. In the case of the CM comparisons, the subjects would perceive the second signal as part of the more complex first signal, while in the sequence MC the additional fundamental in the second signal might be perceived as something 'new', even across different intervals. In the MS and SM cases, such an argument cannot be used because the signals had no component in common. The only common ground to compare both signals is the 'missing fundamental' or – as it was expressed by the subjects – the recomputation of an interval via an octave. In these cases, I assume that the perception of the missing fundamental (or the computation) takes some time. If the complex signal with the missing fundamental (M) is presented before the sine tones, the subject has already some time to 'compute' the fundamental. If the other way around, this extra time is not available and computation starts later. Consequently, the reaction time becomes longer.

4. DISCUSSION

The presented experiments indicate strongly that the case of the missing fundamental is by no means a simple and straightforward auditory task. Musically experienced subjects show a considerably different behaviour in their decision about musical intervals for sine tones, complex signals with a fundamental and complex signals without a fundamental. Especially complex signals with a missing fundamental are either perceived an octave higher, or their perception takes considerably longer than the perception of sine tones or complex signals. In this light, handling the perception of signals with a missing fundamental as equivalent to the handling of signals with a fundamental does not seem appropriate. It is likely that other signals used in psychoacoustic research (e.g.

inharmonic complexes) might show a similar behaviour and are subject to further research. Consequently, some findings for pitch perception might be related to musical perception instead.

5. REFERENCES

- [1] A. J. M. Houtsma and J. L. Goldstein, "The central origin of the pitch of complex tones: evidence from musical interval recognition", *JASA*, Vol. 5, pp. 520-529, 1972.
- [2] E. Terhardt, "Zur Tonhöhenwahrnehmung von Klängen. II. Ein Funktionsschema" *Acoustica*, Vol. 26, pp. 187-199, 1972
- [3] F. L. Wightman, "The pattern-transformation model of pitch", *JASA*, Vol. 54, pp. 407-416, 1973.
- [4] F.A. Bilsen, "Pitch of dichotically delayed noise and its possible spectral basis" *JASA*, Vol. 55, pp. 292-296, 1974.
- [5] T. Houtgast, "Subharmonic pitches of a pure tone at low S/N ratio" *JASA*, Vol. 60, pp. 405-409, 1976.
- [6] F. L. Wightman, "Pitch and stimulus fine structure", *JASA*, Vol. 54, pp. 397-406, 1973.
- [7] H. Davis, S. R. Silverman, and D. R. McAuliffe, "Some observations on pitch and frequency", *JASA*, Vol. 23, pp. 40-42, 1951.
- [8] H. Reetz, "Pitch Perception in Speech: A Time Domain Approach", IFOTT 26, Foris, Dordrecht, 1996.
- [9] A. Seebeck, "Beobachtungen über einige Bedingungen der Entstehungen von Tönen", *Annalen der Physik und Chemie*, Vol. 53, pp. 417-436, 1841.
- [10] J. F. Schouten, "The residue, a new component in subjective sound analysis", *Proc. Koninklijke Nederlandse Akademie van Wetenschappen*, Vol. 43, pp. 356-365, 1940.
- [11] B. C. J. Moore, "An Introduction to the Psychology of Hearing" (3rd ed.), Academic Press, London, 1994.
- [12] R. Plomp, "Pitch of complex tones", *JASA*, Vol. 41, pp. 1526-1533, 1967.