Translingual Discourse in Ethnomusicology 3
2017, 50-84. doi:10.17440/tde017

Psychology of auditory phenomena

ERICH M[ORITZ] V[ON] HORNBOSTEL

Originally published as “Psychologie der Gehörserscheinungen“. In: *Handbuch der normalen und pathologischen Physiologie* 11, 1926, 701-730.

Draft translation: Julia Schöllauf, copy-editing: Jessica Sloan-Leitner

Summarizing texts

Helmholtz, Hermann v.: *Die Lehre von den Tonempfindungen*, 6th edition, 1913 (*Tonempfindungen*). – Stumpf, Carl: *Tonpsychologie*, Vol. 1 1883; Vol. 2 1890 (*Tonpsychologie* I, II). – Stumpf, Carl: *Beiträge zur Akustik und Musikwissenschaft* 1898ff. (9 issues up to now) – Köhler, Wolfgang: “Akustische Untersuchungen I.” (*Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 54:241ff.; *Beiträge zur Akustik und Musikwissenschaft* 4:134ff.); II (*Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 58:59ff.; *Beiträge zur Akustik und Musikwissenschaft* 6:1f.); III and IV. “Vorläufige Mitteilungen“ (*Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 64:92ff.); III. (*Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 72:1ff.). – Köhler, Wolfgang: *Tonpsychologie. Handbuch der Neurologie des Ohres*, 1923, 419ff. (ed. by Alexander and Marburg. – Watt, Henry J.: *The Psychology of Sound*. Cambridge 1917.

Perception and sensation, objectivity

We can generally distinguish between two contrary ways in which sensory inputs appear: the presence and nature of things outside of us, and, in contrast, our moods. We refer to the first – objective – way as perception, and to the second – subjective – way as sensation*;* both ways are referred to as phenomena. This contrast is mostly not an either-or situation but a question of “more or less” depending on the field of sensations, the stimulus conditions and our behavior. The fields of sensations form a descending series according to the objectivity of the phenomena, ranging from the sense of sight to the organ senses. In this series, hearing is between the sense of sight and the sense of touch with a strong connection to the latter also in other respects. Due to this middle position, the two contrary modes of appearance are rather balanced in the sense of touch. Therefore, they were noticed first in this context.[1] (They can easily be experienced when stroking the resting hand with the other hand: The groping hand perceives the other, while the groped hand senses.) But this difference is also apparent in hearing:[2] If we behave naturally, we perceive a truck passing by, but we sense the sound of a tuning fork close to one ear. In other cases, we can easily and deliberately switch between the two modes of appearance: to really listen to the violin over there or to give ourselves entirely to the sound. But depending on the kind of sound, this deliberate change will sometimes be easier and sometimes more difficult. Objectivity not only emerges through our actions. Sounds differ in the natural degree of their objectivity; for neutral behavior, they appear a priori more or less material.

A general prerequisite for something appearing material is that it stands out against a background.[3] All special prerequisites can be attributed to this. At the same time there is the necessity to not only consider the characteristics of the part of the overall appearance standing out as an object but also the characteristics of the background against which it stands out. The degree to which it stands out not only grows with the objectivity of the former but also with the non-objectivity of the latter. Thus, two characteristics can be compared regarding their tendencies to appear more as an object (figure) or more as the background. This contrast, which is essential for all fields of sensation and beyond[4], is also relevant from a biological point of view: Only standing out against the background enables perception. The factors facilitating perception create the outer world.

On top of the formative factors is the shift in space and time – the movement. What is shifted appears as a unit and stands out against the – static – background. Not only silence but also steady sound appears as a background, in particular if it is broad, not very characteristic and not too loud. Game can hear the approaching hunter even from the sound of leaves rustling in the wind.

The prerequisite for the perception of movement is spatial hearing. A sharp sound localized outside the head and perceived with both ears appears material, whereas a diffusely spread, intercranial sound perceived with one ear appears as the background.[5]

The phenomenon which has – without losing its homogeneity and completeness – a richer structure and a “higher” design (Koffka)[6] seems more material than the other. The simplicity of counting the parts, which are often only gained by random fragmentation and not directly given in the phenomenon, is not decisive. A melody does not consist of tones, and language not of single sounds – neither for the singer or speaker, nor for the listener; at the utmost for the acoustician or the phonetician. The overall course, the “movement” of the melody, is more material

---

[1] Weber, E. H. 1846. “Tastsinn und Gemeingefühl.” *Wagners Handwörterbuch der Physiologie* (*Ostwalds Klassiker* no. 149).

[2] Werner, H. 1922. “Grundfragen der Intensitätspsychologie.” *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 10, 68f.

[3] Rubin, E. 1921. *Visuell wahrgenommene Figuren*.

[4] Hornbostel, E. M. v. 1922. “Über optische Inversion”. *Psychologische Forschung* 1, 155.

[5] Hornbostel, E. M. v. 1923. “Beobachtungen über ein- und zweiohriges Hören”. *Psychologische Forschung* 4, 68f.

[6] 1923. *Psychologische Forschung* 3, 363.

and, therefore, more haunting than a single sound or even a series or single sounds – like a profile line which is more explicit and memorable than a point or a curve, which is more descriptive than a numerical series. And like a straight line becoming the basis for a curve, the more monotonous course becomes the background of the sharper course: A regularly repeated drum motif provides the background for the vocal line of an Oriental like a painting on patterned wallpaper.

Unlike optical and tactile objects, sound seems mostly moving and vivid (that is why nothing is more difficultly expressed in music than rest), at least as a process and not actually as an object (e.g., thunder). Usually the source is not out of sight and unreachable. Thus, sound in our perception turns into a strange way of behavior of what we see and touch. We realize how much our hearing is involved in the structure of our outer world when we change the environment: In a new apartment, rooms, doors, handles, drawers and light switches sound strange. An acoustically inclined person will remember such voices for years as essential and personal characteristics of objects and situations. A language emphasizing the essential is abundant with onomatopoeic names, especially the language of primitive people and children (peekaboo, choo-choo; kink). And there are good reasons to assume that all languages were originally the natural motoric-acoustic expression of non-acoustic perceptions and conditions as well. This means that they illustrated all the inner and outer occurrences that affected people through phonetic gestures. Thus, language makes audible what cannot be perceived, and presents what is distant in space and time. Consequently, hearing is maybe the most vital organ in interaction with other people. (Being deaf is usually more burdensome than being blind.) But also for the higher animal species, voice and hearing are more important from a biological point of view than is widely assumed. And also lower animal species, even if they do not "hear" in the strict sense, probably perceive mechanical vibrations, providing them with useful information.

"Sensations of vibrations" are not negligible (e.g., feeling the roughness or smoothness of tangible objects), even for humans. Phenomenally and functionally, these sensations are particularly close to hearing. We can assume the vibration sense to be a preliminary stage in the evolution of hearing.[7]

Noise and tone

We directly perceive a phenomenon only as "this one" or "one of that kind". But we can regard it from different perspectives, compare it to others in different respects and, thus, find out its various "characteristics". This is, again, sometimes easier and sometimes more difficult: Different phenomena have different characteristics which are obvious even without looking for them and become the material core, providing the basis for all other characteristics. This formation regarding substance and attribute is not at our discretion – in extreme cases it is forced upon us, in others it is suggested. It, however, not only depends on the stimulus conditions

[7] Katz, D. 1925. "Der Aufbau der Tastwelt." *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 11, 187ff.; cf. *Jahresbericht über die gesamte Physiologie*, 1922, 377.

but also on the individual predisposition and the particular behavior. Its general, natural direction is often already indicated linguistically by the distinction between noun and adjective: We talk about a bright, sharp and short noise and not about noisy brightness, sharpness and shortness.

The different types of sound – noise and tone – appear so different that a particular receptive organ for each of them has been claimed repeatedly. (Although this assumption is obsolete, only contrasting modes of appearance similar to different senses could lead to it.) But yet, one characteristic does not exclude the other one. Transitions continuously connect the extremes. Of course, it is difficult to produce a tone without any noise and apparently there is no noise a careful "musical" observer would consider completely "tone-free". A more or less regular waveform has been proved to be a stimulus condition, with the wavelength being more important than the amplitude.[8] This means that also from a physical point of view there is no jump but a continuous transition. It leads from pure noise via the vowels and (musical) sounds to pure tones produced by simple sine waves.

Recently, it has been doubted if the sine-shaped form of the wave is maintained in the (peripheral) hearing organ or if it assumes the shape of a sound wave.[9] But even if partial oscillations added physiologically to each fundamental oscillation, the border cases of the series of phenomena would correspond to the sine waves.

We cannot hold a vowel for a longer period of time without changing from speaking to singing. On the other hand, we can only sing vowels and semivowels. If a tone is considerably shortened, it almost completely loses its musical character. (Particularly gifted people, however, may be able to still identify the pitch through a single fundamental frequency oscillation.[10]) So, if we wanted to separate the noises from the musical phenomena, this line would have to be drawn where the frequency is becoming constant and the physical process stationary; but this is never a sharp distinction. Even pure high-frequency sine waves rather produce noise-like phenomena (*s, f*). In very characteristic noises, the tonal component becomes clearly apparent as soon as we destroy partial oscillations through interference.[11] (Nearly) all noises can also be dispersed in this way. Thus, from a physical point of view, noises prove to be similar in nature to musical tones. The two kinds of sound apparently differ regarding their structure: For tone-like phenomena, the waveform has to be simpler and smoother, and for noise-like ones, richer and sharper. Regarded as phenomena, tones and sounds are also quieter, and noises more agitated. The former are more subjective, the latter more objective. From a biological point of view, almost only the perception of noises is relevant. Sounds are rather rare in nature; pure tones hardly ever happen.

---

[8] Weiss, O. and R. Sokolowski. 1920. "Die physikalischen Grundlagen der Geräuschwahrnehmung." *Pflügers Archiv für die gesamte Physiologie* 180, 96–110.

[9] Wegel, R. L. and C. E. Lanc. 1924. "The Auditory Masking etc." *Physical Review* 23(2), 266–285. – Fletcher, H. 1923. "Physical Measurements of Audition." *Journal of the Franklin Institute* 196, 310ff.

[10] Abraham, O. 1919. "Zur Akustik des Knalles." *Annalen der Physik* 60(4), 70ff. – Kucharski, P. 1924. "La sensation tonale etc." *L´année psychologique* 24, 151.

[11] Köhler, W.: "Akustische Untersuchungen III", 85.

Sound color

In general, we can say that from a physical point of view it is the waveform that determines the types of sound and the enormous variety of their characteristics – their sound color (in a broader sense). It is often overlooked that there is, already from a physical perspective, always a single, homogenous waveform which is only divided up artificially during analysis. Two directions are possible: transverse sections, separating parts which follow one another and are intrinsically rather constant; a phonetician uses this method to isolate speech sounds. The other option is longitudinal sections where the “complex” waveform is divided into “partial oscillations”. In both cases, something new is created physically and phenomenally. In the first case, the particularly characteristic transitions and time relations drop out. (The sound of the phonographed sentence “Was it a cat I saw” changes significantly when the glyphs are listened to from back to front.) In the second case we get – through resonance or interference – simple (sine) waves that only emerge because of the test assembly[12] and reunite into the original wave when they all together meet a strongly dampening vibratory entity (air, ear). The same stimulus – a particularly formed wave – and the same phenomenon – a particular sound color – can be technically produced in two different ways: by a single, natural sound source (voice, instrument) or by an artificial synthesis of sine waves of certain frequencies and amplitudes. Only in this sense can we speak of complex waves or compound sounds, and of components or part tones. But each wave per se is physical; each sound color is phenomenally simple, homogenous and without parts. Our consciousness usually knows as little about possible analysis sand synthesis as the air.

Not all types of sound are equally homogenous. We can filter partials from a chord or – with even more difficulty – from a single sound. But what we filter is always weaker than the (technically) isolated component. From a physiological point of view, a part of the “component” energy must remain bound to the overall process determining the color of the sound (or the chord) (see below p. xxx).[13]

Artificial analysis and synthesis enable a correlation of the sound colors with their stimulus conditions. This was recently done for speech sounds, but also for instrumental sounds in broad experimental studies.[14] It has been shown that the color not only depends on the characteristics (frequency and amplitude) of the partial oscillations but also on the particular composition and the overall structure.

---

12 Köhler, W.: *Tonpsychologie*, 432.

13 Eberhardt, M. 1922. “Phänomenale Höhe und Stärke von Teiltönen.” *Psychologische Forschung* 2, 349ff.

14 Stumpf, C. 1918. “Struktur der Vokale.” *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, 333. Berlin;1919. “Analyse geflüsterter Vokale.” *Passow-Schäfer* 12, 234.; 1921. “Tonlage der Konsonanten etc.” *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, 636. Berlin; 1921. “Analyse der Konsonanten.” *Passow-Schäfer* 17, 151. – Miller, D. C. 1916. *The Science of Musical Sounds* – Trendelenburg, F. 1924. “Objektive Klangaufzeichnung.” *Zeitschrift für technische Physik* 5, 236; 1924. “Zur Physik der Klänge.” *Naturwissenschaften* 12, 661 – Crandall, I. B. and C. F. Sacla. 1924. “Dynamic Study of the Vowel Sounds.” *Bell System Technical Journal* 3, 232 – Crandell, I. B. 1925. “The Sounds of Speech.” Ibidem 4, 586.

Thus, the frequencies of the most essential partial oscillations for vocal characteristics (main formant) are not rigid but shift to a particular frequency range with the tone of the voice. Depending on the frequency of the tone, the interval it forms together with the formant is different. The formant usually consists of a series of adjacent partials with a particular distribution of intensity. For whispers, the characteristic components steadily cover a rather broad frequency range. Therefore, a synthesis of sine-shaped partial oscillations and constant wavelength will not be successful. Certain sounds like nasal ones are characterized by a gap in the series of partials; these sounds appear "hollow". Briefly: The structural formula of the different waveforms is given through the energy distribution over the frequency range. It determines the sound color.

Two other aspects have to be considered: First, the ear notices differences in the sound color which the structural analysis – at least with the means available so far – overlooks. Improved methods might enable further development. But it has to be taken into consideration that even the most accurate structural formula gained through analysis does not provide a totally adequate description of the physiological – and maybe even physical – form corresponding to the sound color. This description suggests that although the detected "elements" build a certain constellation in the synthesis, they remain unchanged concerning their fundamental characteristics. But this is certainly only the case to a limited degree. Synthetic sound can rather be compared with a chemical compound than with an arrangement of independent parts. In some cases, a sound color may appear similar to another one that "contains" the structural formula as elements – å "has something" of o and a; violet has something of red and blue in it. In other cases, the whole is something entirely new which cannot be compared to its artificially isolable parts. In the first impression of the sound of a French horn, the partials have disappeared like the spectral colors in white. Even in a two-tone chord of a very low and a very high tone, the tones do not appear completely separated from one another, but the higher one seems in a way embedded in the lower one. In general, higher components are more or less hidden by lower ones.[15] A hearing impaired person who was able to hear isolated high tones from the range of the e-formant very well, nevertheless perceived a synthetic e as an ou.[16]

Second, so far structural formulas have only been developed by analyzing single sounds. It is widely known that sentences are easier to understand than words and words easier than nonsense syllables. Of course, we add missing information, improve subjectively, catch the meaning of the well-known and listen into it; beginning and ending, length and rhythm, sequence and connection of the sounds are at least as characteristic as the sounds themselves. In this way, even musicians cannot recognize the timbres of instruments anymore if beginning and ending are missing. Sung vowels are easier to identify if they are preceded by a

[15] Watt, H. J.: *The Psychology of Sound*, 62. – Wegel, R. L. and C. E. Lane: "The Auditory Masking & c. 1924. " *Physical Review* 23(2), 266–285.

[16] Claus, G. 1923. *Beiträge zur Anatomie, Physiologie, Pathologie und Therapie des Ohres, der Nase und des Halses* 19, 294–304.

consonant.[17] But even more: Destroying the high formants makes the isolated single sounds s, t, i, l unrecognizable but not the word "still"[18]. It blurs i but not mi.[19] The spoken vowel with its shifting frequencies is more characteristic than a vowel sung on one constant tone. If sounds merge, they themselves obviously change, too – even objectively through the way they are produced. The overall form is more than a sequence of the same partial forms characterized by structural formulas.

Brightness[20]

If we play a record on a gramophone in a faster or slower mode, what we hear will change in a certain respect, no matter what we have recorded – music, speech, noise: It will be brighter in one case and darker in the other. In one direction, this change appears as an increase and based on this increase the entire characteristic is called "brightness" (and not "darkness").

In Middle High German, the term "hell" – bright – was only used in its original acoustic meaning. The fact that is was completely transferred to the optical field is a strong argument for the identity of the intended phenomenon in both fields. In fact, we can easily and precisely find a shade of grey as bright as any given tone. Different observers will come to equal matches. (Protanomals – and dark adopted – require much darker, and deuteranomals require much brighter shades of grey than the normal ones.) Smells, for instance, can also be associated with equally bright tones or shades of grey.

Thus, brightness refers to the phenomena of several, if not all fields of sensations. From an evolutionary point of view, it is one of the oldest characteristics of phenomena (and the physiological processes) and, therefore, the most resilient one against disruptions of the normal process. If the length of the sound is significantly shortened, or sometimes also if partial oscillations (and, thus, the timbre) are destroyed, brightness remains the only characteristic of the phenomenon. At the extreme ends of the audible frequency range we can only identify brightness. In pathological cases, the musical character of tones and sounds can completely disappear; the sound colors can fade. But brightness does not seem to drop out.

This is similar in the optical field: At the ends of the visible spectrum we can only identify differences in light intensity, not in shade. Shortening the stimulus duration makes chromatic colors achromatic. Some blind people notice a "glimmer" if they turn to light or dark. There exist colorblind people, but there are no people with color vision who are blind for brightness.

---

[17] Stumpf, C. 1918. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften*, 343, 357. Berlin.

[18] Stumpf, C. 1921. *Beiträge zur Anatomie, Physiologie, Pathologie und Therapie des Ohres, der Nase und des Halses* 17, 186f.

[19] Stewart, G. W. 1923. *Physical Review* 21, 718.

[20] Bibliography in Stumpf: *Beiträge zur Akustik und Musikwissenschaft* 8, 17f., 21f.

Movement. Pitch. Distance

An increase of acoustic brightness is not only an increase in a static phenomenon, like brightening up a viewed color field, but a rise such as the upward movement of a visible objet. The impression of movement and its direction are so stringent that most languages use the terms "rise" and "fall" instead of "high" and "low".

It is by no means justified to limit this linguistic usage to tones because the phenomena of all types of sound are equal and equally strong in this respect. "Movement of a melody" has the same meaning in speech and music. The up and down corresponds to the immediate, natural impression and has not added only associatively through "spatial symbolism". The hand movements accompanying the singing of all "primitives", the movements of dancers and conductors – as far as they are not limited by rhythm and conventions – follow involuntarily (therefore very much in hypnosis)[21] the movement of the melody. The course of the movement – besides rhythm – is decisive for the shape of the melody and its expressive meaning, and is an essential carrier of meaning in speech. Differences of dialects mainly refer to the speech melody. Objectively continuous change, like already in single sounds in speech and in the glissando of a singer, corresponds to the visible "real" movement. But also stroboscopic "quasi" movement has its precise analogy in the step changes of brightness from one syllable to the other, from one tone to the next. In both cases, this movement is even stronger than a continuous shift. There as here, the movement becomes more vivid (within a certain range) if the interval between the stimuli becomes longer – legato, staccato – and the distance increases – step, jump.

The span of the movement (length of the step, distance) is determined rather precisely: For a step given as a pattern, we can find another one of equal length originating from another (third) level of brightness, particularly for tones but roughly also for noises.

For noises, however, accuracy cannot be measured because physically the frequencies determining the brightness cannot be clearly quantified.

On the other hand, we consider tones equal on the basis of the overall appearance (interval, see below). It seems that the distance cannot be clearly isolated.

A static chord also appears the broader the further apart the frequencies of the limiting tones are. The strong, noise-like buzzing of a number of reeds of adjacent frequencies can be compared extremely precisely concerning their width, although the limiting tones (and in general partials) cannot be identified. The phenomenal width only depends on the frequency ratio of the limiting tones and not on the number and adjacencies of the components.[22] The sound width also contributes to the characteristics of noises (e.g. consonants).

[21] v. Schrenck-Notzing. 1904. *Die Traumtänzerin Magdeleine C.*, 132f.

[22] Abraham, O. and E. M. v. Hornbostel. 1925. "Zur Psychologie der Tondistanz." *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 98, 233.

Expansion. Weight. Density

In order to describe the change of the phenomenon with the stimulus frequency, a third contrastive pair besides bright–dark and high–low is suitable: small–big. A rise is a contraction at the same time; a fall appears as an expansion. This is also an immediate acoustic impression and does not require sensations of other senses. Young children spontaneously speak about small things in a high voice and about big ones in a low voice. The meaning of a phonetically equal word changes, for instance, in Sudanic languages because of "high tone" and "low tone" in the same sense (small–big).

Psychologists name the expansion of acoustic phenomena as "volume" in most cases. Compared to space, sound is in fact more three-dimensional than flat or linear. However, the expansion of a single sound is not at all spatial in the proper sense, for example, noise at a city square or humming on a meadow. Furthermore, "expansion" alone does not completely describe the intended aspect of a phenomenon. Big sounds are leaky, labile, loose, diffuse, soft, dull, and moreover ponderous, viscous, inflexible, heavy; whereas small sounds are dense, solid, compact, concentrated, hard, pointed, and at the same time vivid, flexible, light. (The Greek refer to this contrast as βαρύς and όξύς.) The terms point to a mass that seems heavier and more inertial for darker sound (weight) and to its distribution within the expansion, resulting in higher density and solidity for smaller sound. At the same time, which one of the two aspects is more apparent depends on the volume: The weight predominates for loud and low as well as soft and high sounds, and density for loud and high as well as soft and low sounds.

The same assignment of characteristics is obvious in, for example, sine curves. It is certainly not by chance that these characteristics – in their physical sense – also apply to sound sources and waves. But they have not been transferred from there to the phenomena. We do not have to know anything about the volume of a pipe and the wavelength, but nevertheless we hear the dimension of the tone.

Also in optics, dark colors seem heavy, and bright ones light; dark colors appear spatial (deep, without resistance against immersion), and bright ones like surfaces (solid). These aspects of phenomena are obviously closely bound to the brightness which can be found in all fields of sensations. Together with this brightness, the phenomena form the – phylogenetically oldest and most stable – core of the phenomena which is essentially decisive for the ratio between object and background (see above p. 51).[23]

The question if there are also other "volumes" besides brightness does not make a sense from a pure phenomenal point of view because there is no doubt that we mean different things if we refer to bright, small or dense tones. Only through the relationship to the "dimensions" of the physiological and, further, the physical processes the modes of appearance change with, can we determine that the possibility of somehow independent change should decide for the assumption of a particular characteristic. The sensitivity for differences in the frequency of simple

[23] Gelb, A. 1920. "Wegfall der Wahrnehmung von ‚Oberflächenfarben'." *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 84, 193.

tones is higher if the change in brightness is observed instead of the change in "volume".[24] A less spread excitation on the basilar membrane may correspond to frequent stimuli, be it that we assume a sharper tuning of the shorter fibers,[25] a reduced bulge of the membrane[26] or a shorter extension of the oscillation processes from the oval window upwards[27]. But deliberately isolating the factors could be even more difficult – it is never entirely successful. Comparing it with the stimulus magnitudes might not show significant differences, and corresponding peculiarities of the physiological or physical processes might not have been discovered yet. Nevertheless, the description would not have to be in accordance with it and the variety of the observed phenomena would not be less real. All factors are bound to one another to a certain degree: "Bright" somehow also implies "solid", and "solid" also "bright". They can probably only by separated when the senses are increasingly differentiated in their development while here the one side and there the other become particularly apparent in the homogenous entity ("bright–solid"; bright for the hearing, solid for the pressure sense).

Vocality[28]

Regarded as phenomena, spoken vowels are in the middle between noises and musical sounds. The distinctly characterizing aspect of the phenomenon is more or less apparent in all types of sound – already in noises and still in simple tones. "Already" and "still" are to be understood in an evolutionary sense because vocality undoubtedly belongs to the older characteristics of sound closer to the core, though not belonging to the core itself like brightness. If the sound is very short, vocality can disappear; it is also missing in some noises. For simple tones, it becomes less clear from the ends towards the middle of the frequency range (minimum at about 2000 v.d.[*]). The similarity between the tones and the vowels constantly changes with rising frequency. It goes through the series of m (about 132 v.d.)–u–o–a–e–i–s (about 8200 v.d.) with all the nuances in between.

Not all vocalities appear in speech. The human vocal tract cannot produce some of them – the ones between m and u as well as i and s. (As a consequence, the vocalities are not transferred into the tones but identified in them.) On the other hand, the vocalities emerging vocally in the continuous transition between o and e via umlaut ö are missing for simple tones – like the purple shades in the color spectrum.

Within the continuum, the indicated vocalities mark a change of the direction of similarity, like the "unique hues" in the shaded color scale: At the transition

24 Rich, G. J. 1919. "Tonal Attributes." *American Journal of Psychology* 30, 121.
25 Waetzmann, E. 1912. *Folia neuro-biologica* 9:24.
26 Wegel, R. L. and C. E. Lane. 1924. The Auditory Masking & c. *Physical Review* 23(2), 266–285.
27 Watt, H.: *Psychology of Sound*, 162ff.
28 Köhler, W.: "Akustische Untersuchungen II, III"; *Tonpsychologie* – Stumpf, C. 1914. *VI. Kongress für experimentelle Psychologie* (*Beiträge* 8, 17).
* [*vibrations doubles*, i.e., c.p.s. (or Hz). Eds.]

from å to ä, the similarity to o in a pure a disappears and the similarity to e starts, like in the transition from violet to orange where the blue shade in pure red turns into a yellow shade.

The intervals between pure vocalities are approximately octaves.[29] In trying to determine the turning points, like for the unique hues,[30] we face peculiar difficulties. This also strongly depends on the method. In both fields, the aim can only be reached if we arrange the stimuli in – graded, ascending and descending – series and not randomly.[31] The weak manifestation of vocality in tones makes this task more difficult, in particular for musical people. But also other people have to get used to the requested perspective of judgement, and practice sticking to it. It seems that we not only have to disregard Tonigkeit (see below) and brightness but also other dimensions of vocality itself: Tones, especially those not totally free from overtones, seem u-like up to the two-line octave and then, up to the four-line octave, ü-like.[32] Whether the series from u to i via ü or the one via o–a–e is perceived, and if a certain frequency, for example, 1700 v.d., is recognized as ä or ü, depends on external and internal conditions (that have so far not been investigated in detail). In any case, conditions can be found under which the vocalities (also o, a, e) cannot only be observed even for simple tones, but also frequency ranges spanning an interval from a half to a whole tone can be found in which the turning points are in the middle.

Like the unique hues which can be produced by blending the lights they lie between – for example, unique yellow from reddish yellow and yellowish green – also the pure vocalities arise from the connection of in-between frequencies, for example, pure o at the consonance of tones that independently would resemble ao and uo.[33] Indeed, vocality seems more manifest in such connections than in simple tones,[34] like sung and spoken vowels that are always characterized by a range of formants. It is not a single frequency acting as a formant but a frequency range[35] (as a consequence, the vowels are in the middle between tones and noises also regarding their function). In very complex sounds, according to the less clearly defined frequency, the brightness (and in some cases also the Tonigkeit, see below) is not so clearly determined. It becomes more often the neutral background on which color (as a "figure") becomes apparent. At the same time, however, the vocality "valences" that do not fit together – in the mentioned example the a- and u-similarities – seem to neutralize each other (like complementary colors). Thus, the physiological processes that determine vocality can completely destroy each other in waves of a certain structure, leading to noise without vocality. This noise only

[29] Also in R. König. 1882. *Quelques expériences d'acoustique*, 42ff.

[30] Westphal, H. 1910. *Zeitschrift für Sinnesphysiologie* 44:182.

[31] Rich: *American Journal of Psychology* 30, 131ff. – Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 8, 43ff.

[32] Köhler, W.: "Akustische Untersuchungen II", 126. – Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 8, 44f.

[33] Köhler, W.: "Akustische Untersuchungen III/IV", 97.

[34] Köhler, W.: "*Akustische Untersuchungen III*", 33 (S by two Galton whistles).

[35] Cf. also E. Jaensch. 1913. *Zeitschrift für Sinnesphysiologie* 47, 219.

assumes vocality again if we delete certain stimulus components (through interference).[36]

According to their resemblance, the vocalities can be arranged in a scheme like colors. It has to be considered that u and o as well as i and e differ less than o and a as well as e and a. U and o are related to i and e in a way similar to the relation between the “cold” colors (blue and green) and the “warm” ones (yellow and red). This fact, together with the already mentioned one that u–ü–i are easier to perceive in sounds with less characteristic vocality (like simple tones), suggests that this series may be phylogenetically older than the o–ö–e series – like blue–yellow is older than green–red. This fits to the fact that the character of the vowels approaches the one of noises the more distant they are from a, which is, so to say, the purest vowel in this respect. But at the same time it has the vaguest, flattest and less characteristic vocality.

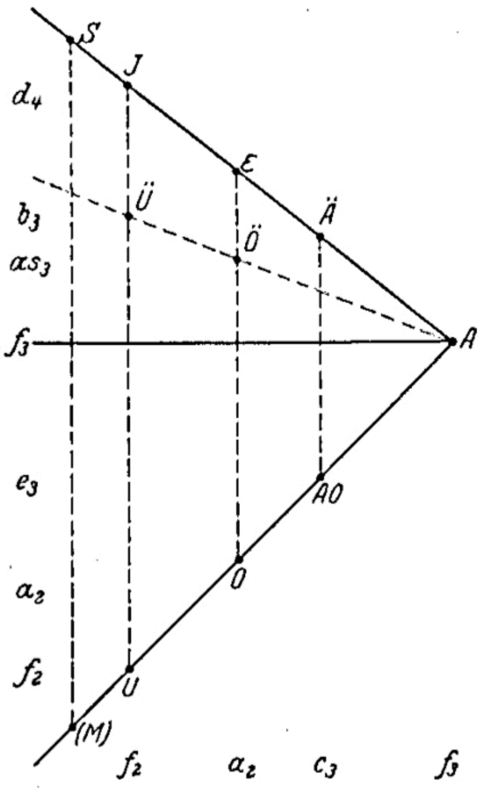


Fig. 1. Vowel triangle

The arrangement of the vowels in the scheme according to their mode of appearance leads to a triangle with a at the top (fig. 1). If we draw the level curve originating from this point horizontally,[37] brightness is descriptively illustrated, too.

[36] Köhler, W.: “Akustische Untersuchungen III”, 85.
[37] Also in Chr. Fr. Hellwag. 1781 (reprint ed. by Victor, Heilbronn 1886).

The ratios of the figure are based on the frequencies of the formants of whispers (according to Stumpf) – the upper formants in the ordinate, and the lower ones in the abscissa (increasing from left to right). The main vowels with their transitions are on the sides of the triangle. Sound continua with equal lower formants are on vertical sections, and sound continua with equal brightness and (from right to left) increasing characteristic are on horizontal sections; an expansion beyond the u–i basis already leads to "semivowels". Not all sounds within the triangle can probably be found in human speech. Some can be found in animals: the darker sounds are found in big animals (mammals, e.g. cattle), while brighter sounds are found in small animals, in particular birds. For the latter, for instance, transition sounds from s to m, which are usually hard to imagine, can be observed.

Chroma [Tonigkeit]

For most people, if we change from a spoken vowel (e.g. a) to a sung one and further to a tuning fork sound (of equal pitch, e.g. C_0), vocality retreats more and more into the background and another characteristic becomes more apparent: the characteristic that is the reason for why we compare "musical" types of sounds with all other sorts. (Even the primitives prize singing above everyday language and see it as most effective means of cult and magic.) Let us call it "Tonigkeit" [chroma] (see p. 84).

Together with the stimulus frequency, Tonigkeit also changes, but in a different way than brightness or vocality. For a given tone, tones can be found in other registers that are very similar apart from brightness – much more similar than closely adjacent tones. The frequency ratio of the thus found tones is 1:2:4 …. They form "octaves".

A whistled C_2 sounds darker than a C_2 played on a piano and even darker than a sung C_1. Even very "musical" people cannot escape this impression.[38] Tonigkeit remains the same, in the first case strictly – it depends on the fundamental period of the wave. Brightness, on the contrary, changes with the timbre – the sound structure. In their chants with small descending intervals, the Papuans from the Torres Strait make an octave leap upwards as soon as they reach the lower end of their voice range, which obviously does not appear as an interruption of the melody but only as a (technically unavoidable) change of register.[39] Children transpose up by one octave when repeating tones in their own register as soon as they generally start to repeat tones (under one year of age).[40] But only if they are "musical" – if not, they try to repeat the brightness and timbre of the pattern as precisely as possible, regardless of Tonigkeit. Adult non-musical people react in the same way.[41] (Therefore, repeating tones is a good method to

---

[38] Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 8, 23.

[39] Myers, Ch. S. 1912. *Reports of the Cambridge Expedition to Torres Straits* 4, – v. Hornbostel. 1913. *Jahrbuch der Musikbibliothek Peters* 19:18ff.

[40] Stumpf. C.: *Tonpsychologie* I, 293; *Beiträge zur Akustik und Musikwissenschaft* 8, 27.

[41] Köhler, W.: "Akustische Untersuchungen III", 56.

prove musical talent.) Animals seem to totally lack awareness of octaves. If dogs are trained to react to a certain sound when food is provided, it is not more difficult to break them from reacting to the octave of the desired sound than to any other frequency.[42] They are, however, very sensitive to different timbres – like non-musical people. Consequently, the “food sound” played on another, even very similar instrument has no effect.

There is no stringent proof that animals do not perceive Tonigkeit in phenomena, but it is very likely. They obviously perceive only one overall characteristic, with no separable aspects, in what they hear. Therefore, parrots and other birds learn to repeat whistled melodies at the original pitch level, which is always objectively implied. They do not “transpose”. But this does not mean that their phenomena have Tonigkeit because they also change the timbre of the model as little as they can. Warblers often imitate noises deceptively. When they repeat something, parrots also try to copy the register (brightness). This means that they react like non-musical people who repeat tones. Even the “best” singers mix the sounds we perceive as very tonical with various tone-free noises as if they did not see any substantial difference between the two.

Tonigkeit strongly decreases in very high, very low and very short tones, even for musical people – it may even disappear completely.[43] It can be regarded as the most unstable, biologically least important and evolutionary youngest factor of acoustic phenomena.

Nevertheless, it is of peculiar interest for psychologists – not only as the basis of music. It is a very strange characteristic which lets octave tones appear so similar that we might consider them equal in this respect, but at the same time it does not change as continuously in a certain direction as in a continuous transition from one tone to its octave – like brightness or vocality. If we compare a tone to another one a few frequencies higher and if we completely ignore brightness, etc., both appear equal (regarding Tonigkeit). If we boost the second tone a few more frequencies, it suddenly appears completely different from the first one (still regardless of brightness, etc.). The same happens if the reference tone approaches the octave: Initially, it is totally different from the first tone, but all of a sudden it becomes as similar as possible (or even identical).

In Rich’s[44] experiments, in which he tried to measure the sensitivity to differences in the various factors of tonal phenomena, it proved difficult to find the suitable question for Tonigkeit. In the end, a musical subject had the idea to ask, regardless of brightness, if “the two tones were identical or not”. The thus found thresholds were significantly higher than the ones for brightness, but they were still much lower than the smallest intervals applied in music (about $^{1}/_{10}$ of a half-tone).

The position of the turning point or the width of the range where the tones seem identical strongly depends on the test conditions, in particular on the ob-

---

[42] Pfungst, O. (oral report); cf. C. Stumpf: *Beiträge zur Akustik und Musikwissenschaft* 8:55.

[43] Köhler, W.: “Akustische Untersuchungen III”, 25, 42, 45.

[44] Rich, G. J. 1919. *American Journal of Psychology* 30, 159.

server’s attitude. If the difference between the stimuli increases continuously, the observer can stick to the “equal” perception as long as possible and only ask, “Is it still going on?” This also happens involuntarily if the tone changes in small steps or even continuously (the stimulus “sneaks in”). The same applies for the “unequal” perception of descending tones. But the observer can also lie in wait for the difference and ask, “Has it already changed?” For medium range differences between stimuli, we can voluntarily let the perception switch. However, the width of the range in which this is possible differs according to the observer and his current condition. The same applies for optical spatial figures, for example, at the transition from ᴍ via ᵛ to ∨.

We cannot watch a continuous series of flashes, neither hear or perform a continuous series of knocks without structuring it involuntarily and often unwittingly. The way we structure it does not matter – whether we “accentuate” certain flashes or knocks and diminish other ones, or (subjectively) lengthen, brighten up or strengthen them. Most of the time, all these changes occur at once. But it is essential that the various single knocks or lights form little groups, and that these little groups then form larger groups. It all has to form into a structure which is easier to grasp. The simplest, most natural structure is (2 x 2) x 2, etc. (This also applies to natural processes such as egg cleavage.) Let us now assume that the *phenomenon Tonigkeit is based on such a structuring form of in principle regularly periodical central-physiological processes.*

It follows the facts of binaural hearing that the central processes are periodical and their frequencies exactly match the ones of the stimuli, but they cannot be superposed like oscillations.[45]

“In principle” requires an explanation: One option is to imagine that the structure does not start immediately at the beginning of the process. In this case, *Tonigkeit* would become apparent only after a short, maybe quantifiable latency period. If the stimulus duration is significantly shortened, *Tonigkeit* is also very weak or does not appear at all. Corresponding “Gestaltzeiten” [latency until the gestalt effect sets in] have also been found in other fields, for example, in stereoscopic depth perception.[46]

Like in heard oscillations, and also in groped ones, irregular (noise-like) oscillations can be distinguished from regularly periodic (tone-like) ones. Again, the periodical oscillations have a longer latency period.[47]

Alternatively, the structure could be effected only in higher centers, for example, in the cortex, whereas subcortically the process would remain regular. Thus, it might become comprehensible why *Tonigkeit* is less apparent or not existent at all in animals and non-musical people. In any case, a physiological change even without our voluntary action will be necessary so that *Tonigkeit* is constituted as directly as other shapes if the individual disposition is sufficient and the conditions are not too unfavorable. We could only make assumptions about the *kind* of

[45] 1923. *Psychologische Forschung* 4, 72ff.
[46] Karpinska, L. v. 1910. *Zeitschrift für Psychologie* 57, 1.
[47] Katz, D. 1925. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 11, 208.

changes when a clearer image of the processes in the acoustic sector had been developed.

Provided there is a structure, the sound seems tonical according to our assumption. It is essential that the process be periodical: The more regular the periodicity – in this sense, already the "simpler" the wave form of the stimulus – the easier the structure. Therefore, *Tonigkeit* is less apparent in noises. In "complex" waveforms, the fundamental period obviously has to determine the structure and, thus, *Tonigkeit*. This is why we hear sounds (if we do not filter "partials") in the *Tonigkeit* of the fundamental oscillation, even if it is weak or at all missing[48] (see below p. 72). And this is why the main voice in the parallel fifths and fourths of the Middle Ages and of contemporary indigenous peoples is the lower voice.

But *how* is the periodical process structured? First, according to the multiples of two, as we must conclude from observable analogous cases. How far this fragmentation goes will depend on the frequency and the "physiological time of presentness";[49] that is, on how many single periods form an overall process which does not fragment, such as when we involuntarily form groups of 2, 4 or 8 depending on the tempo as we count beats. This might also be related to the fact that *Tonigkeit* is less apparent at very low or very high frequencies. For very low frequencies, too small and hardly structured groups would fall into the time of presentness, and for very high frequencies, too big and very structured groups would fall into this same time. Also in such cases – and also for noises – *Tonigkeit* becomes immediately more apparent in a musical context. This means that the preliminary concise structures make the structure easier in the following ones – through "perseveration" or "attitude". *Tonigkeit* will not change at the same given frequency, no matter how far (to which power) the structure goes, if only the principle of the structure (in this case according to 2^n) remains unchanged. The *grade* of the structure will be decisive for the manifestation of *Tonigkeit*, the *kind* of structure for the quality. A sudden single tone also seems tonical, always – every time from the beginning – in the same *Tonigkeit*. In this case, the most natural structure (according to 2^n) will occur.

Tonal relatedness. Interval

Isolated single tones are, however, the exception. Sound sequences are the rule. The theory of *Tonigkeit* has to prove itself primarily for the rule. The first tone already establishes a certain structuring principle – it lays the ground, so to say, for the following one. If the following tone is an octave of the first, the structure remains the same regarding its nature. Only the number of the sub-elements that are combined into one element changes. This element is as long as the former regarding time. Therefore, also the *Tonigkeit* remains "the same", despite the change in brightness. This is generally valid: Even if the first tone has not been structured according to the multiples of 2 but of 3, 5, etc. (which will be rather common in a

[48] Köhler, W.: "Akustische Untersuchungen III", 123ff.
[49] 1923. *Psychologische Forschung* 4, 120f.

broader musical context), the structuring principle at the transition to an octave remains the same.

Octaves. Structure according to 2 x 2^n

8	⓪	•	○	•	ʘ	•	○	•	⓪	•	○	•	ʘ	•	○	•	⓪
4	ʘ		•		○		•		ʘ		•		○		•		ʘ
2	○				•				○				•				○
1	•								•								•

Octaves. Structure according to 3 x 2n

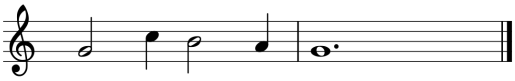
4	⓪	•	•	○	•	•	ʘ	•	•	○	•	•	⓪
2	ʘ		•		•		○		•		•		ʘ
1	○				•				•				○

We can easily illustrate the rhythmic patterns – most simply in rows of dots which can indicate at the same time, for instance, the transition through the zero point or other equidistant phases of the physiological or physical process. The rows can be extended to any dimension. Further, the starting and closing elements are not relevant, even if the homologous elements of two rows are shifted against each other (phase is without influence). Only the overall structure is relevant: We can also understand it by gradually regarding the different rows and consider it similar or dissimilar.

The structure according to 3 comes closest to the one according to 2 regarding simplicity. This distinguishes (besides octaves) fifths and fourths among all ratios of *Tonigkeit* – as steps, motif frames and the basis for transposition in monophonic music.[50] (The equivalent for rhythm is the – equally smooth – transition from duplets or quadruplets to triplets or vice versa.) Regarding *Tonigkeit*, also a simple tone seems similar to its fifths and fourths, though not identical as to its octaves. For the appearance of *steps* of fifths and fourths, it is not irrelevant where the interval starts from, apart from the changes in brightness and pitch: Rising fifths and descending fourths usually seem to diverge from the normal level, whereas descending fifths and rising fourths seem to reach the goal or come to rest. An example is the sound sequence *C*-*G*-C_1 (or vice versa) with the structures 2 x 2^n, 3 x 2^n, 4 x 2^n. The main structure (2^n) remains the same; only the structure of the smallest sub-elements changes. The structure of *G* has been prepared by the one of the preceding *C* (or C_1; G in "*C*-color", G^c). (Following the linguistic usage in music, the structure according to 2^n shall be named "tonical"; the tone with a structure determining the structure of other tones [including its octaves] in a context shall be named "tonic", and the resulting structural relation itself "tonality".) This is different for the sequence *G*-C_1-G_1, if ***C***[51] has not already been determined as the "tonic" with a defined structure (according to 2 x 2^n) from the musical context. If ***G*** is structured according to 2 x 2^n, the structure changes significantly at the transition to ***C*** – even more significantly the more fragmented ***G*** is (the higher *n* is), that is the more manifest the ***G***-*Tonigkeit* is. (For $n = 0$, the smallest group to

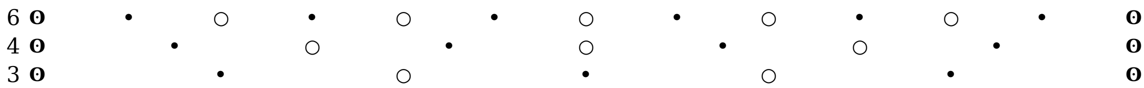
[50] v. Hornbostel. 1913. *Jahrbuch der Musikbibliothek Peters* 19, 21ff.

[51] According to Stumpf, tones are labelled with Fraktur letters, regardless of their octave. [This has not been maintained in the English version. They are set in bold letters. Eds.]

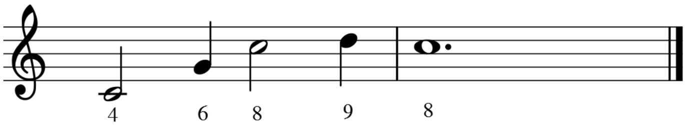
be changed has already six elements, for $n = 1$ it has 12 elements, etc.) As the structure of ***G*** does not provide a preparatory determination, ***C*** will also be structured "tonically" (according to 2^n) – the step will thus appear as a "change of the tonic". But this slightly changes the perception of ***G*** retrospectively, as if it had not been the tonic but the under-fourth ("dominant") like in the first case. But if the musical context permanently lays the focus clearly on ***G***, ***C*** appears as a modulation and tension which requires a return to the initial point, for example, in

(This is obviously related to the question of why musicians often regard the fourth as "dissonant". It has to be emphasized that we only refer to a melodic context at this point which must not be derived from knowledge of harmony or interpreted through imagined harmonies.)

Fifth above, fourth below the tonic

Fourth above, fifth below the tonic



In the pure melodics of the peoples who do not know harmony, all intervals notably narrower than a fourth are used without a distinction. They are melodic "steps" κατ' ἐξοχήν [kat' exochín, "par excellence"], only distinguished according to their width (distance) which is determined by the shape of the melody, but most of the time, as already noted by Helmholtz[52], they are variable within a wide range. They all seem tonal, also the "thirds", like a transition to an opposed adjacent tone and thus as a progress, not as a change of register like in octaves, fifths and fourths. In fact, the ratios 4 : 5, 5 : 6 and 6 : 7 are not predominant over 7 : 8, 8 : 9, 9 : 10 etc. regarding their restructuring. Even 5 and 7 are obviously instable as entire groups – they are difficult to identify rhythmically and phenomenally. Therefore, this structure will hardly occur and easily change into a simpler one (in particular the tonical one according to 2^n). (In some contexts, 8 : 9 could be more obvious than the other seconds, for example, in



[52] v. Helmholtz: *Tonempfindungen*, 6th ed., 422f.

Furthermore, there is again the threshold: The more complicated the structures are and the more similar the distances become, the less different the intervals become – also physiologically.

In every sequence of two tones there is a quantitative factor – the length of the step or the distance – and a qualitative one – the "interval color" – at the same time; both depend on the frequency ratio of the stimuli. Thus, this ratio determines the general phenomenon "interval" (see above p. 57). The distance changes continuously together with the frequency ratio (like brightness, pitch, etc.), while the interval color does not (like *Tonigkeit*). Some interval qualities stand out a priori, but distances do not. The sensitivity to differences in distance is significantly higher than in interval colors. If the stimulus duration is shortened, the interval color[53] becomes less apparent or disappears completely; the distance remains. This means that the distance needs a certain period of time to develop. Also in this respect, distance and interval color behave like brightness and *Tonigkeit*. Like the distance which is based on a change in the level of brightness, we assume that the interval color is based on a structural change forming the physiological correlate of *Tonigkeit*. Depending on whether this change is smoother or more sudden, we name tones between which this change happens related or unrelated. This *relatedness, however, is neither determined only by the ratio of the frequencies, nor only by the ratio of the structures: Both ratios have to fit together.* Different structural conditions may correlate with the same frequency range, depending on the circumstances, like the above-mentioned example of 3 : 4 has shown. The same example has also shown that, even if the structuring principle (2^n) remains the same, the tones may seem unrelated. They only seem related if the structural change required by the frequency ratio is adapted to the already existing (field) structure. Provided that they fit together, we can deduce some rules to determine the degree of relatedness:

1. The structures with the same and homogenous structuring principle, that is structures according to 2^n or $3^{n,}$ etc., must be nearest related.

Within such "pure rows" it seems appropriate to also graduate according to the degree of structure (n) and, for instance, to consider a tone more closely related to its octave (like mother and child) than to its double octave (like grandmother and grandchild). Consequently, octaves would only be very similar, also regarding their *Tonigkeit*, but not identical. This approach is certainly justified, in particular regarding the row $3^0 : 3^1 : 3^2$, but hardly relevant in practice because the quickly increasing distance reduces the similarity of the tones and intervals. Furthermore, the degree of the structure is not so clearly determined for a single tone (see above). This means that it may in fact remain unchanged at the transition to the octave (e.g., $2 \times 2^2 : 4 \times 2^2$ instead of $2 \times 2^2 : 2 \times 2^3$).

2. Structures are the more closely related the lower the prime numbers determining the structure are. $2^n : 3^n \gtrsim$ (more closely related than) $2^n : 5^n$; $2 \times 2^n : 3 \times 2^n \gtrsim 2' \times 2^n : 5 \times 2^n$. It can be assumed that the structuring prime numbers do not exceed 7.

53 Likewise the consonance of two-tone chords. Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 4, 24.

3. The higher structure is more relevant than the lower one: $3 \times 2^1 : 5 \times 2^1 > 2 \times 3^1 : 2 \times 5^1$; $3 \times 2 \times 2^1 : 2 \times 3 \times 2^1 \gtrsim 2 \times 3 \times 2^1 : 2 \times 2 \times 3^1$. (This theorem is particularly suitable to explain the essence of the structural theory in contrast to other ones which are only based on numerical proportions. In the last example, *Tonigkeit* "per se" is always the same. Which of the different structures occurs only depends on the context.)

Consonance

The phenomena in chords show some new characteristics compared to single tones and sound sequences. Among them, the phenomenon of "consonance", which was a problem already for the ancient Greek, has attracted most interest and led to detailed theoretical discussions and experimental investigations due to its particular significance for the music of Europe since the end of the Middle Ages (see below p. 77). The fundamental question from a phenomenal perspective is as follows: What is the factor according to which we arrange two-tone chords in a row ranging from the octave on one end to "seconds" and "sevenths" on the other end?

We directly perceive a chord in its overall appearance. Only a comparison, while examining the basis for this comparison as precisely as possible, leads to the different "sides" of the phenomenon; they lead to more or less different arrangements of the row.[54] Which arrangement is the "correct" one? Which side is relevant for consonance?

Three factors do not come into question in the first place: First, pleasantness and other emotive impacts which vary from era to era, from culture to culture and from individual to individual. Second, raucousness which can – induced by beats – be added or removed (by spreading the tones on both ears) artificially.[55] Third, sound width (see above p. 57) which leads to a totally different arrangement of the row and can vary considerably in two-tone chords of nearly equal consonance degree (e.g. octave–double octave, thirds–sixths, seconds–sevenths).

What remains can probably be described best as tones *fitting together* to a more or lesser extent; extreme mismatch [can] also [be] positively [described] as conflict. (symphony = consonance [literally "to sound together" from German "Zusammenklingen". Eds.] , diaphony = dissonance [literally "to sound apart" from German "Auseinanderklingen". Eds.] in ancient Greek; harmony {from ἁρμόττειν} to fit together, already in Plato).[56] From a phenomenal perspective, fitting together *as such* only occurs if the sound in a way appears as a – even very homogenous – complex. For the rest, only the *effect* of fitting together, which can be characterized as homogeneity, simplicity, balance, unity or the opposites, becomes apparent

54 Malmberg, C. F. 1918. "Perception of Consonance and Dissonance." *Psychological Monographs* 25(2):93 – Pratt, C. C. 1921. "Some Qualitative Aspects of Bitonal Complexes." *American Journal of Psychology* 32, 490.

55 Stumpf, C. 1898. "Konsonanz und Dissonanz." *Beiträge zur Akustik und Musikwissenschaft* 1.

56 Stumpf, C. 1897. "Geschichte des Konsonanzbegriffes." *Abhandlung der Bayerischen Akademie der Wissenschaften* I (21), 13f.

in a phenomenon. In many cases, the way of acting can be changed deliberately while the stimuli are constant. It is possible to switch between observing the match of the complex components and the homogeneity, etc. of the unanalyzed overall phenomenon. The fact that its degrees are in accordance under both conditions leads to the conviction that both modes of appearance have the same basis.

In sound sequences, we can correspondingly either observe how smooth a transition is in the homogenous phenomenon “interval” or if two tones are related – and fit together – directly in the phenomenon “pair of tones”. But here, the degrees of the two modes of appearance will not always be in accordance because the second isolating behavior (which therefore has not been mentioned above) might create different conditions, reduce or delete the influence of the precedent, etc. *This* difference is of course irrelevant for a (single) chord.

The hypothesis developed for *Tonigkeit* and tonal relatedness can now easily be applied to chords. Matching structures of the central-physiological processes will correspond to the phenomenal match,[57] and the homogeneity of the physiological overall structure will match to the homogeneity of the overall appearance.

We observe something analogous to consonance also in vibration sensations: If we softly and simultaneously touch the tines of the vibrating tuning forks 55 and 110 with two adjacent fingertips, our impression is very different from that of tuning forks 55 and 100. The first impression, compared to the second, can only be described as closer harmony, greater homogeneity, unity, softness – briefly as consonance. Involuntarily, it can be precisely identified. An octave sounding consecutively (possibly without interruption) appears as a smoother transition, but the impression is no longer as clear as for consonance. We cannot observe any similarity or dissimilarity (apart from the difference in brightness) in isolated stimuli with longer pauses between them. However, everybody is probably “non-musical” at first regarding vibration sensations.

Initially, we have to empirically determine the sequence in which the two-tone chords are arranged according to their consonance degree. We face several difficulties in this process, particularly due to the effect of the three above-mentioned factors which do not come into question.

Not only musically gifted persons have to be careful not to be influenced by their musical experience and the consequential emotive values even in laboratory experiments. This applies even more to so-called non-musical people because they may be influenced unconsciously. The rate of how often they mistake a two-tone chord for a unison provides an objective measure of homogeneity, which can be gained particularly from non-musical subjects (Stumpf); however, subtle differences cannot be detected in this way. The raucousness of beats makes sounds appear inhomogeneous, even without any theoretical knowledge. Alike, separated localization left and right reduces the homogeneity of the immediate impression. (Systematic experiments with dichotic sounds are still to be carried out.) Undoubtedly, homogeneity decreases with increasing sound width. It is, therefore, often difficult to decide what in fact causes phenomenal inhomogeneity; this is particu-

[57] Similar in Köhler: “Akustische Untersuchungen III”, 131.

larly the case for two-tone chords larger than one octave.[58] Besides these difficulties, there are others: Homogeneity increases if one stimulus is weakened – chords continuously pass into sounds. It changes with the timbre.[59] Thus, solid insights concerning the consonance degrees cannot be gained from observing homogeneity only. It has to be directly investigated if the tones of the complex fit together; this must be ultimately decisive.

Experimenters and theorists roughly agree regarding the row of the consonance degrees of two-tone chords (within an octave): octave, fifth, fourth, thirds and sixths, tritone 5 : 7, sevenths and seconds. They also agree about the consonance in this row being continuously decreasing and the dissonance being increasing. These terms name contrasts such as white and black, but not two mutually exclusive categories. The contrasts are linked through intermediate stages, though not continuously like the ends of a grayscale. A dividing line would be arbitrary.

We may sum up the intermediate stages and form a third (transition) group, name these stages in a particular way (Krüger's "neutral sonances",[60] Watt's "paraphonies"[61]), and point out their characteristics. But even then the division remains arbitrary, does the fourth belong to the consonances and the tritone 5 : 7 to the dissonances, or do both belong to the neutral ones? By the same token, others could claim neutral zones between the three groups, etc. There would be endless border conflicts. The gap between some levels (e.g., between octave and fifth) is certainly wider than between other ones (e.g., between a major and a minor sixth), but an estimation is even difficult for these clear differences and completely impossible for various "seconds" and "thirds", for example, which differ regarding their consonance degree at best and could only be integrated into the row according to the latter. If anything, we could separate the two-tone chords, which can in some way be (directly) compared and put in an order regarding their consonance degree, from all the other ones. The two-tone chords of the first group would stand out because the tones somehow "sound together" – the structures fit together. Consequently, we would have to regard them as consonant in this broader sense. The "dissonances" would be identical to the – musically "useless" – consonances that are "out of tune". This disjunction would at least be grounded in the phenomena, even if it does not correspond to the linguistic usage.

The more detailed order of thirds and sixths, seconds and sevenths and the "intervals of seven" between these two groups (seventh 4 : 7, tritone 6 : 7, third 6 : 7 and their "inversions") differs, depending on the observation conditions and theoretical assumptions. This means that initially the structure hypothesis only has to be in accordance with the solid approximate arrangement. It can be easily deduced from the above-stated rules for structural relatedness if we consider that the

[58] Stumpf, C. 1898. "Neueres über Tonverschmelzung." *Beiträge zur Akustik und Musikwissenschaft* 2, 14ff.

[59] Stumpf, C. 1898. "Zum Einfluß der Klangfarbe auf die Analyse von Zusammenklängen." *Beiträge zur Akustik und Musikwissenschaft* 2, 168.

[60] Krüger, F. 1908. *Wundts Psychologische Studien* 1, 305ff.; 2, 205ff.; 4, 201ff.; 1910. 5, 294ff.

[61] Watt, H. J. 1919. *The Foundations of Music*, 155ff.

main structure in a chord is provided by the joint fundamental period (according to the third rule).

The fundamental period can, but does not have to be, given objectively through the lowest tone of the chord or a difference tone. If there are such difference tones, they emphasize the structure and thus contribute to the *manifestation* of the consonance. But they can be missing without changing the nature of the structure and the consonance *degree* determined by it.[62]

The same also applies to the overtones of sounds which Helmholtz considered determining for consonance phenomena. *If* they are there and coincide – this means they are louder – they increase the conciseness of the overall structure and, thus, the stability, unity and homogeneity of the phenomenon, but not the consonance degree. (This is as if the dots in our figures were more blackened or otherwise emphasized, but were the *same* dots that had already been underlined.)

A corresponding relation even exists between simple tones and sounds, which are, from a physical point of view, nothing but chords with a particular energy distribution: The structure in sounds is a given objective. *Tonigkeit* is clearer in sharp and rich sounds than in pure tones and even clearer in (unanalyzed) octave chords. People whose "absolute tone awareness" it not very reliable can easily recognize the "key" of a piece, a monophonic melody or a major triad, even if they are in the dark about single sounds, in particular those with few overtones.

In instrumental sounds, the fundamental tone is often very weak – from a physical perspective and, if we can filter it out, also from a phenomenal point of view. (In this case, *Tonigkeit* often seems moved down by octaves[63].) The structure of the sounds is based on the joint fundamental period, even if it only exists in its own multiples – physically and physiologically.

The subtle differences regarding the consonance degree as they result from the rules for structural relatedness are not so clear from a phenomenal perspective, but at least they do not oppose the observations made so far. According to the structure, the major third 4 : 5, for example, is more consonant than its inversion, the minor sixth 5 : 8, because in the third the main structure is simpler as opposed to a sub-structure (according to 2^n) in the sixth (rule 3). For the same reason, the minor third 5 : 6 is less consonant than the major sixth 3 : 5. Rule 2 states further that 4 : 5 is more consonant than 3 : 5, and 5 : 8 more consonant than 5 : 6. Further, 3 : 5 is more consonant than 5 : 8 (rules 3 and 2). This results in a (descending) row 4 : 5, 3 : 5, 5 : 8, 5 : 6, which is in accordance with the row empirically found by Pear[64] and Malmberg[65]. (The question in Malmberg's test series concerned "blending"; the definition "a seeming to belong together, to agree" is in

---

62 Krüger, F. 1903 "Differenztöne und Konsonanz." *Archiv für die gesamte Psychologie* 1, 205ff.; 2, 1ff.; "Theorie der Konsonanz." *Wundts Psychologische Studien*, see footnote 60. – Stumpf, *Beiträge zur Akustik und Musikwissenschaft* 4, 90ff.; 5, 1ff.; 6, 151ff.

63 Köhler, W.: "Akustische Untersuchungen III", 128f.

64 Pear, T. H. 1911. "Differences between major and minor chords." *British Journal of Psychology* 4, 56.

65 Malmberg, C. F. 1918. "Perception of Consonance and Dissonance." *Psychological Monographs* 25(2), 93 – Pratt, C. C. 1921. "Some Qualitative Aspects of Bitonal Complexes." *American Journal of Psychology* 32, 490.

accordance with my definition of consonance as fitting together.) If the seventh 4 : 7 is perceived as more consonant than the minor sixth 5 : 8,[66] this also becomes understandable (from rule 3).

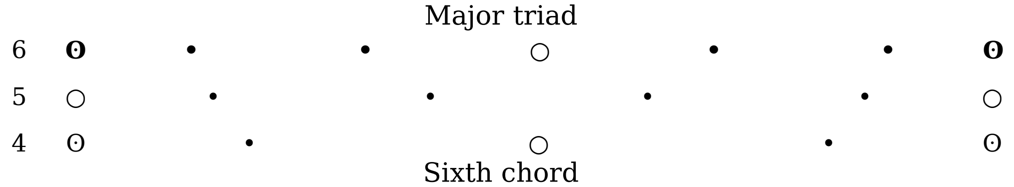
Theoretically (according to rule 2), the second 8 : 9 could indeed be more consonant than the third 4 : 5, or even more consonant than the fourth 3 : 4 (according to rule 3). Systematic experiments – in which the two-tone chords extended by octaves would have to be compared in order to avoid beats – have not been carried out so far, but occasional observations seem to *support* the assumption. Thus, also parallel seconds would become understandable, which are, besides parallel fourths and fifths, common in duets of some peoples.

Like for two-tone chords, the same approach can also be applied to three-tone and multiple-tone chords. Following what has been explained above, consonance in them will be more manifest in most cases. But the match of everything involved – phenomenally all tones, and theoretically all structures – determines the consonance degree. This means that it is always determined by the overall structure and cannot be deduced from the consonance degrees of the two-tone chords. The three-tone chords 2 : 3 : 4 and 3 : 4 : 6 both “comprise” octave, fifth and fourth; both only comprise structures according to 2^n and 3^n. Nevertheless, the first is more consonant because the main structure (2^n) is simpler.

If we consider the *Tonigkeit* of octave tones identical and not only closest related (see p. 68f.), we must consequently assume the same consonance degree for a two-tone chord and its octave extensions. (This can, however, hardly be verified empirically.)[67] Thus, octave doublings in multiple-tone chords are irrelevant for the consonance degree (but neither for conciseness, nor for homogeneity).

According to its overall structure, the major triad 4 : 5 : 6 is (theoretically) more consonant than its inversions – phenomenally it is at least more stable. The six-four chord [second inversion] 3 : 4 : 5 is superior to the sixth chord 5 :6 : 8 in the sense that the structuring numbers regularly rise with the frequency, and the lowest structure thus refers to the fundamental tone. However, the simplest structure (2^n) is hidden in the middle – which makes the six-four chord appear particularly instable – whereas in the sixth chord the simplest structure is in the highest and most exposed tone. In the root position of the triad, the two outer notes share a sub-structure according to 2, and in the sixth chord the two upper ones share this sub-structure. In the six-four chord it is probably completely missing (because the overall structure becomes simpler). These structural characteristics will also be relevant for the stability of the overall appearance.

Major triad



Sixth chord

[66] Preyer, W. 1879. *Akustische Untersuchungen*, 64. – Krüger, F. 1903. *Archiv für die gesamte Psychologie* 1, 219.

[67] Stumpf, C. 1898. *Beiträge zur Akustik und Musikwissenschaft* 1, 78ff.; 2, 14ff.

8 Θ • ○ • Θ • ○ • Θ

6 Θ • • ○ • • Θ

5 ○ • • • • Θ

Six-four chord

5 ○ • • • • ○

4 ○ • • • ○

3 ○ • • ○

For the minor triad 10 : 12 : 15 and its inversions, the situation is far more complex. Only *one* conclusion from the structural theory shall be explained. Like in sound sequences, different structural ratios at equal frequency ratios are also possible in chords. Thus, the fundamental tone of the minor triad 10 can be structured according to 2×5^1 or 5×2^1, and the triad 12 can be structured according to $2 \times 2 \times 3^1$, 3×2^2 or $2 \times 3 \times 2^1$, etc. On the one hand, the actual structures will depend on the constellation of all tones involved, including the difference tones. On the other hand, they will depend on the musical context: the precedent structures (when performing music in practice, also the subsequent, mentally anticipated structures). Within the ***E***-tonality (e.g., at the change from major to minor) and for the triad *E-G-B*, the structure of *E* (5×2^1) and *B* (5×3^1) will be determining. In the ***G***-tonality it will be the structure of *G* (3×2^2) and *B* (3×5^1). The structure of the third tone will adapt to the overall structure as much as possible.

15 Θ • • • • ○ • • • • ○ • • • • Θ

12 Θ • • • ○ • • • ○ • • • Θ

10 Θ • • • • ○ • • • • Θ

15 Θ • • ○ • • ○ • • ○ • • ○ • • Θ

12 Θ • • ○ • • ○ • • ○ • • Θ

10 Θ • ○ • ○ • ○ • ○ • Θ

Besides the adaption of a sub-structure to an already existing field structure, also two structures may adapt to each other. Such adaptations can be observed in different kinds of phenomena and have partly already been investigated, for example, in optics. Maybe they will lead to a general theory of the threshold in the future. In any case, they explain that the consonance degree remains unchanged if there are small deviations from the simple frequency ratios, although from a physical perspective the most complex vibration ratios emerge particularly here. But as it has already been mentioned, the central processes are not waves and their superposition. Consequently, there are no central beats; phase shifts are irrelevant for the timbre of sounds and the consonance of chords, like for the relatedness of consecutive tones (see above p. 66). The adaptation of the structures does not at all require a central change of the *frequencies*. The tones 200 and 401 obviously show a good octave consonance because the structures 2^n and 2^{n+1} still fit well together in them and, as we can observe, even better than in 200 and 400.

In order to appear ideal, octaves (and other consonances) have to be augmented a little in a chord and significantly in a sequence. The required detuning of the simple frequency ratio increases with the augmentation of the interval and rising frequencies.[68] This totally unexplained fact so far may have its parallel in the "overestimation of filled distances": A row of dots seems to be longer than the distance between the isolated endpoints.

*   *   *************

A series of knocks, at an objectively constant tempo, also seems to be slower in sixteenth notes than in quarter notes. We tend to increase the tempo together with the fragmentation. Likewise, the stronger fragmentation of the higher octaves may seem to prolong the period and lower the octave tone and, thus, result in a narrowed interval; even more the more strongly connected the tones are – the higher their relatedness or consonance.[69]

It depends on the stability of the structural connection if we notice small detuning: The threshold is higher – which means that the sensitivity for detuning is lower – for simultaneous tone pairs than for consecutive ones[70] (like in other fields); it is higher for sounds rich with overtones than for simple tones[71]; it is higher for more perfect consonance than for little consonance[69]; and for three-tone and multiple-tone chords, it is higher than for two-tone chords.

Like in sound sequences, the factor based on brightness (sound width, see above p. 57) in chords can hardly be separated from the factor based on *Tonigkeit* (consonance). Therefore, the thresholds are always determined by both – a separation has not been aimed for in experiments so far. We can only assume that width – like distance in sound sequences and brightness in single tones – is a more precise criterion and involuntarily preferred if we ask for the subtlest differences. We can further only assume that the width has an even purer effect the less manifest the consonance is. This is probably the main reason for the mentioned different thresholds. But there is also a third factor: musical experience. The determination of certain tones on instruments, resulting in a practical preference of certain intervals (apart from octaves, fifths and fourths), leads to denominations[72] ("*C*", "major third"). "Absolute tone awareness" and "interval awareness" are such denominations. We can further investigate to what extent an "*A*" of 435 oscillations can be

68 Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 2, 125ff. – Maltzew, C. v. 1913. *Zeitschrift für Psychologie* 64, 213ff. – Köhler, W.: "Akustische Untersuchungen III", 3f.

69 Stumpf, C. and M. Meyer: *Beiträge zur Akustik und Musikwissenschaft* 2, 130.

70 Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 1, 55; 2, 129.

71 Stumpf, C. and M. Meyer: *Beiträge zur Akustik und Musikwissenschaft* 2, 131.

72 This does certainly not refer to logical categories but to facts of one's consciousness. A psychological analysis of the latter cannot be provided at this point.

raised without being perceived as an "$A^{\#}$" and to what extent a "minor third" in a sequence or a chord can be augmented without being perceived as "major". Such "historical qualities" (Stumpf) prove to be extremely flexible and strongly dependent on the experimental design, the attitude and, of course, the individual conditions. (For example, we might perceive an objective major third as a "minor" one and a tritone still as a "fourth".)

Even in experiments focusing on other problems, musical observers will hardly be able to free themselves from their habits. The perfect fifth 2 : 3, for example, is the best fifth for a violinist.[73]

We can certainly increase the number of terms defined by experience and practice. If recent attempts to expand our tonal system to a 24-step system are successful, new tones like "*+C*" and new intervals like a "neutral third" will be part of the musical awareness and limit the scope of the old terms.

The names of tones and intervals facilitate their identification, but do not enable it. Somebody could, for instance, know the keyboard of a piano without the names of tones and intervals, and show on the keys what has been sung.

Another mode of appearance of chords is linked to musical practice and, thus, to the changing fate and history of cultures. The order of the consonance degrees has remained the same since ancient times and would, as far as it could be tested, probably prove to be the same for all people. The *relative* simplicity of the structures is a purely scientific fact, like the order of the tones according to their brightness. This is different for the *absolute* simplicity of the structures: What is easy to understand for a well-prepared person is confusing, chaotic and incomprehensible for others. What is rich for one person seems vapid for others. In this sense, octaves and fifths have become empty, flat, poor and insubstantial in modern Europe. For sons, a dissonance, which was still spicy for their fathers, is for them gentle. These modes of appearance cannot be dissolved into the emotive impacts they might have. Also the youngest spread light and shadow, softness and hardness, and sweetness and bitterness in their art. But they need a different segment of the row which is equally graduated for all. Our ancestors would not appreciate our softest sounds even if their tastes required *more* spiciness than ours.

Historical criticism

A detailed discussion of the consonance theories of even only the past half century would go too far at this point. The theory outlined so far has a lot in common with earlier approaches. If we emphasize them and their main characteristics, the basis the theory builds on will become more apparent. The only hypothesis of the theory resulted from a Gestalt theoretical[74] discussion of the specific musical characteris-

[73] Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 2, 127.

[74] Köhler, W. 1920. *Die psychischen Gestalten*. – Wertheimer, M. 1922. *Psychologische Forschung* 1, 47ff.; 1923. 4, 301ff.; 1925. *Drei Abhandlungen zur Gestalttheorie* – Koffka, K. 1922. *Psychological Bulletin* 19, 531ff., 1925. *Psychologie im Lehrbuch der Philosophie* (ed. by. M. Dessoir).

tic of acoustic phenomena – *Tonigkeit*. A theory of sound sequences and, further, of chords followed easily. Due to this evolution, difficulties and concerns became obsolete which, because of the priority of monophonic music, arise from the theories based on consonance phenomena. Even Stumpf[75] could not solve these concerns. The fundamental hypothesis integrates acoustic phenomena into the wide range of psychological, physiological and physical facts which have become more understandable regarding their nature and their regularities by applying the structural concept. Thus, also problems which can hardly or not at all be solved experimentally in the field of acoustics may become comprehensible through investigations in other fields.

1. The structural concept of consonance is most closely related to Stumpf's *Verschmelzungstheorie* [theory of fusion]. The concept of "fitting together" seems to be nearly identical to the concept of "*Verschmelzung*" which has, however, often been misunderstood. (In order to not increase the confusion, I have avoided the term so far.) Stumpf defined *Verschmelzung* "as the connection of two sensations, forming a whole or uniformity, as an approximation of a two-tone chord to monophony"[76]. He considers it the essential fundament of consonance, but he also underlines repeatedly that we (directly) perceive consonance *only* where the consonant and fused tones are clearly separated. It is an *effect* of fusion and not fusion itself that it is difficult or – for non-musical people – impossible to distinguish them. Fusion can serve as a means to indirectly determine the consonance degrees (see above p. 70). But as discriminability according to Stumpf also depends on several other factors – difference in brightness, volume ratio, etc. – it is not as clear a criterion as fitting together. It is comprehensible that the two may produce different results.[77] Fusion does not directly depend on the frequency ratio of the stimuli but on the parallel processes of the phenomena in the brain.[78] Ascribing (phenomenal and functional) fusion to the relevant modifications of the central-physiological processes is a postulation for which analogies from adjacent fields can be useful.[79]

This is the starting point for the structural hypothesis. Conclusions drawn from it lead in several directions beyond the scope of the *Verschmelzungstheorie*, particularly where the latter has caused concerns (which could partly be solved together with the misunderstandings they had emerged from).

Thus, the consonance phenomenon is not limited to two-tone chords because also three or more tones can fit together better or worse, sound more or less together, and form a more or less homogenous entity. A triad is not equal to the sum of three two-tone chords with the same consonance degrees they would have on their own. It is, on the contrary, a new entity with its own consonance degree and can be compared with other chords in this respect. The consonance degree of a triad is not the result, for example, the average of the consonance degrees of the

[75] Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 1, 55ff.

[76] Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 1, 44.

[77] Malmberg, C. F. 1918. "Perception of Consonance and Dissonance", *Psychological Monographs* 25(2), 93 – Pratt, C. C. 1921. "Some Qualitative Aspects of Bitonal Complexes." *American Journal of Psychology* 32, 490.

[78] Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 6, 120.

[79] Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 1, 50ff.

two-tone chords we can analytically filter out of it (and to a certain degree prove regarding their own consonance): The triad *C–E–*A^b in tempered tuning ($A^b = G^{\#}$) is not only more dissonant than the major third (*C–E, E–*$G^{\#}$) and the minor sixth (*C–*A^b) if we refer to musical experience, but also a pure and immediate phenomenon.[80] Phenomenally, the consonance degree of *C–E* does not necessarily remain unchanged when $G^{\#}$ is added[81] because the structure of *C–E* can change under the influence of the new conditions – also physiologically. It is more likely to change the more favorable the conditions are for such an influence. These environmental conditions are mainly the musical context. The sudden changes of the mode of appearance it causes (e.g., at "enharmonic changes") must also have physiological reasons.

Strictly speaking, adding or deleting tones always creates new conditions. In this sense, also overtones and combination tones have an influence on the consonance degree.[82] We cannot generally say that a two-tone chord of simple tones changes if it is combined with others, nor that it remains unchanged, but the two-tone chord becomes a multiple-tone chord. Its consonance is determined by all structures involved and their fitting into *one* overall entity.

The fact that different structures at equal physical conditions are possible is also the basis for different possible "perceptions", even in a melodic sense. The context and the musical experience support a certain perception – in some cases they even enforce it. Thus, isolation also supports a certain structure (e.g., for tones the "tonical" structure according to 2^n). We have to strictly distinguish between the structural change according to the *nature* and the structural change according to the *conciseness*, which also depends on the context, but, furthermore, on several other factors that do not influence the nature of the structure (e.g., the consonance degree): absolute and relative volume and length, (extreme) frequency range, timbre (see above p. 72), the distribution of the stimuli to both ears (dichotic hearing), but in particular the predisposition of the listener. Some objections to Stumpf's *Verschmelzungslehre*[83] are settled if we consider the differences between the "dimensions".[84]

2. The structural hypothesis is also linked to the so-called *rhythm theories* of consonance (Lipps[85] and others[86]). It deserves this name even more as is it based

80 Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 6, 139ff.

81 Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 6, 123 passim.

82 Peterson, J. 1925. "A functional view of consonance." *Psychological Review* 32, 17.

83 For example, against the theorem of the irrelevance of the volume; *Beiträge zur Akustik und Musikwissenschaft* 2, 9.

84 Geheimrat Stumpf told me that he himself has changed his tonal theory in 1917. The new version is in some aspects similar to the one explained above, in particular because it abandons the assumption that consonance is based on fusion.

85 Lipps, Th. 1885. *Psychological Studies* 1, 92ff.; 1905. 2, 115ff.; 1902. *Zeitschrift für Psychologie* 27, 225.

86 Cf. Stumpf: *Beiträge zur Akustik und Musikwissenschaft* 1, 23ff. – The theories tracing consonance to simple frequency ratios without being able to provide sound arguments for this connection can be ignored at this point. Cf. Stumpf: *Beiträge zur Akustik und Musikwissenschaft* 1, 19ff. (Leibniz, Euler); Max Meyer. 1900. *Psychological Review* 7,

on time shapes with a hierarchical structure, whereas Lipps (like many others) even regards the regular periodicity of oscillations as rhythm. According to his theory, only the maxima of a wave have a physiological effect. Thus, the central processes are discontinuous; consonance depends on the concurrence of impulses. This means that rhythm in a narrower sense (shaping through differentiation of the elements of a row into more and less accented ones) only occurs in multiple-tone chords. Accents would emerge through the sum of concurring impulses – through amplification. But periodic amplifications can be heard as beats or raucousness if their frequencies are not too high. An octave of (not too low) simple tones, however, is totally smooth. If it is only an imperceptible, unconscious beat, there is still the paradox that the consonance degree increases and decreases with its frequency (according to Helmholtz) like the dissonance degree with the raucousness of the audible beats – even a simple analogy with the phenomena fails. Furthermore, as Stumpf stated, the impulses only concur if the phases coincide. At every phase shift, accents are deleted, resulting in an unstructured row.

Meanwhile, this assumption of discontinuity, which is impracticable also from a physical and physiological point of view, is not necessary for the rhythmical structure; even Lipps gave it up later. It settles the main objections. The basis for consonance is homogeneous overall structures formed of rhythmical sub-structures. This is what Lipps obviously had in mind as the core of his theory. In parts, he found conclusions similar to the ones above resulting from the structural theory, for example, regarding the “tonic”, the “double interpretation” of the fourth, etc.

The reason why the phase shift does not have an influence is that the elements of the (sub-) structures can be confined by random (equal) phases without any change of the structuring principle (p. 66). Consonance is constituted by the sub-structures that fit together as a *whole*. Regarding neural activity, we have to keep in mind that the central processes exist side by side without a superposition (p. 64). They must not start one-sidedly from the physical side – from “outside” – but at the same time from the observable phenomenal side – from “inside”. Then and only then, the point of criticism that it is easy to ascribe arbitrary, wonderful abilities to the unconscious is as unjustified as the one stating that randomly invented “physical images” are put in the place of controllable psychological facts. The assumption of physiological – that means unconscious – structures, which correspond to the phenomenal rhythms, does not exclude differences between the two kinds because even the possible times are of totally different dimensions. Microrhythms *as such* obviously cannot be perceived. But this does not mean that they cannot either exist physiologically or appear in another way (as *Tonigkeit*). Likewise, we do not hear periodical processes and their frequencies but sounds of different brightness. The combination of duplets and triplets may already be hard to comprehend in acoustic, optical and motoric rhythms[87], but corresponding micro-

241; 1901. *University of Missouri Studies* 1, 1; 1903. *American Journal of Psychology* 14, 192.

[87] Stumpf’s objection to Lipps: *Beiträge zur Akustik und Musikwissenschaft* 1, 27. Krüger, F. 1903. *Archiv der gesamten Psychologischen Studien* 1, 96f.

rhythms can result in perfect consonance, maybe just because the elements concentrated in such short periods of time connect even stronger.[88]

Determining the temporal constants of neural processes will probably lead even further. It is at least noticeable that the oscillation period of the lowest audible tone (about 16 [c]ps) coincides with the time (about 60 σ) that constitutes the lower end of (clear) phenomenal succession – obviously in all fields of sensation (succession threshold).

3. Stumpf used compelling arguments to disprove Helmholtz' theory that *beats* are the reason for dissonance phenomena: There are (artificial) beats without dissonance and there is extremely sharp dissonance without beats (in adequately created chords of simple tones, in distributed tuning forks, at diplacusis, in our imagination[89]). If the unison is only slightly out of tune, we hear only *one* tone and can imply the other tones only from the beats. However, it does not sound as homogenous as a single tone but confusing and, in this respect, similar to a dissonant, though beatless chord. Thus, the impression contingent on the multiple sound provides at least an indirect criterion of dissonance similar to the criterion of consonance provided by phenomenal unity, but it is very unreliable and rough.

Consequently, the question of smoothness or raucousness leads to another hierarchy of two-tone chords than the one according to match, fusion, simplicity or unity (not to be analyzed). This hierarchy is also opposed to musical consciousness. Only the extremes – octaves, semitones – remain in their positions, but the fifth might seem more raucous than the major sixth; the minor third belongs entirely to the dissonances.[90]

A related question resulting in a similar row is the one of similarity between the overall impression and a pure tone. This similarity is without a doubt stronger in perfect consonances than in dissonances. “Purity” is, so to say, the phenomenal equivalent (but not a contrast!) of the multiplicity of beating chords. It is also an indirect criterion – an effect of the match, not the match itself – and equally unsuitable for determining consonance: In one case, an overtone-free single tone would “ex definitione” become a perfect consonance. In the other case, a rattling noise would be an extremely sharp dissonance. The consonance degrees, not only their manifestation, would depend on timbre and intensity.

Purity as a phenomenal characteristic is in the focus of all theories which see the nature of consonance – or its essential condition – in the *coincidence of partials* (Helmholtz[91]) or *difference tones* (Krüger[92]), and, relatedly, in being free of beats.

---

[88] Cf. Lipps. 1885. *Psychologische Studien* 1, 96f.

[89] Stumpf, C.: *Beiträge zur Akustik und Musikwissenschaft* 1, 4ff.

[90] Malmberg, C. F. 1918. “Perception of Consonance and Dissonance”. *Psychological Monographs*, 25(2), 93 – Pratt, C. C. 1921. “Some Qualitative Aspects of Bitonal Complexes.” *American Journal of Psychology* 32, 490.

[91] v. Helmholtz: *Tonempfindungen*, 10th section.

[92] Krüger, F. 1900. “Beobachtungen an Zweiklängen.” *Wundts Philosophische Studien* 16, 307ff.; “Theorie der Kombinationstöne.” 1901. Ibidem 17, 186ff.. See also F. Krüger. 1903 “Differenztöne und Konsonanz.” *Archiv für die gesamte Psychologie* 1, 205ff.; 2, 1ff.; 1908. “Theorie der Konsonanz.” *Wundts Psychologische Studien* 1, 305ff.; 2, 205ff.; 4, 201ff.; 1910. 5, 294ff.

The resulting overall sound is the poorer – this means physically closer to the simple tone – the more components the single sounds have in common. This is, in fact, also a sort of fitting together: Adding an octave that is already completely "included" in the fundamental sound does not disturb the existing order. But if we take sounds of stopped [gedackt] pipes which only include uneven numbered partials, all partials of the octave sound will be in the gaps between the partials of the fundamental sound. The twelfth, however, would not add anything new to the fundamental sound; it would be more consonant than the octave, which would – apart from being beatless – belong to the dissonances.[93] We find the same paradox if we trace the similarity of octave sounds to common partial frequencies in a successive comparison with Helmholtz ("relatedness of sound"). At the utmost, the similarity of timbres can be traced to equal elements (though not completely and under the restrictions mentioned above p. 55f.) if the partials with their valences of brightness and vocality contribute to the overall brightness and timbre. But the similarity of octaves has nothing to do with that: The brightness of octave tones as well as of vocalities varies significantly. But the similarity implied in the relatedness of sound refers to *Tonigkeit*. Like consonance, this similarity obviously cannot be traced to equal elements, pointing to the fact that only the structures can be similar. An equilateral triangle does not become more similar to a square just because both have sides of equal length – this means being made from the "same elements". For the same reasons, Helmholtz' theory of the relatedness of sound is disproved together with his theory of consonance. Like for the consonance of octaves, we cannot refer to experience in sounds for the similarity of simple octave tones, even if they strongly prevail. We cannot either refer to the transfer of such experience to the perceptions in simple tones by our "memory". The relatedness of sound further does not emerge only secondarily through perceptions of consonance in chords. Consonance is not the reason for the relatedness, neither vice versa, but both have the same fundament: the structures which cause *Tonigkeit*, the characteristics of sound sequences and chords – in short, all "musical" phenomena of the sense of hearing.

Loudness

If a tuning fork is struck by a falling pendulum, the amplitude of the oscillations of the tines increases with growing drop height. The tone gets *louder* at (nearly) equal timbre, brightness, vocality and *Tonigkeit*. If we let the strongly struck tuning fork fade out, the tone becomes *softer* with decreasing amplitude and disappears completely in the end. This characteristic of the sound phenomenon shares some aspects with brightness: It forms a one-dimensional continuum which appears, passed through in one direction, as an increase; it is parallel to one variable of the stimulus – the amplitude, like brightness is parallel to the frequency; it is not specific to hearing but has its equivalents in the nature of the change of other sensual phenomena we generally refer to as "intensity".

[93] See also v. Helmholtz: [6]*Tonempfindungen*, 346f.

For “lower” senses, apart from thermoception, it is quite clear what we mean by this term, but it is more difficult for the sense of sight. Equating optical strength and brightness, which is still very common, cannot be justified.[94] (Physicists[95], in particular, tend to conclude “equal causes, equal effects”, without considering that sound and light waves, and their amplitudes and frequencies can have very different effects on different organs and still be a long way from the periphery to the cortex.) The comparability of acoustic and optical brightness (p. 56) forces us to distinguish between brightness and strength also in optical phenomena.

Together with loudness, like with brightness, sharpness, etc., the tendency of a sound to stand out against other sounds, push them to the background and become a figure itself increases (p. 51). This functional efficiency – the “level of efficiency” – becomes apparent in the phenomenon as being “haunting”. As this efficiency goes parallel with loudness but does not only depend on it, “pure” loudness can never be observed directly.[96] Thus, the absolute threshold, that is the energy amount of the stimulus at which a sound becomes audible or inaudible, is not lower in binaural than in monaural hearing.[97] If the differences are very small, it is often impossible to decide whether they belong to loudness, brightness or timbre. As a consequence, it is technically hardly possible to fulfill the claim to keep the irrelevant factors constant when measuring the sensitivity to differences.

Further, loudness not only depends on the amplitude of the wave: It varies with the frequency at constant amplitudes. The obvious assumption that the relevant factor is the energy and not the amplitude cannot be carried out because the threshold energy has a minimum at about 3,000 oscillations and increases to both ends of the audible frequency range.[98] The same relation between stimulus energy, frequency and loudness seems to be valid for tones above the threshold.

Recently, McKenzie[99] has overcome the difficulty of qualitatively comparing different tones regarding their loudness in an indirect process which is modeled on flicker photometry: Reference and normal stimuli are presented in quick succession. Focus is on a smooth row.

Accordingly, equal loudness at any frequency corresponds to the same multiple of the threshold energy at this frequency. This theorem can also be deduced from Fechner’s law, stating that loudness is a logarithmical function of the stimulus intensity. Thus, this law has been proved valid in experiments despite all possible objections.

---

[94] Stumpf, C. 1917. “Die Attribute der Gesichtsempfindungen.”, *Philosophisch-historische Klasse* 8. Berlin.

[95] Auerbach, F. 1924. *Tonkunst und bildende Kunst*.

[96] Werner, H. 1922. “Grundfragen der Intensitätspsychologie.” *Zeitschrift für Psychologie* 10, 18ff.

[97] Stumpf, C.: *Tonpsychologie II*, 430ff., in particular 439. – v. Hornbostel. 1923. *Psychologische Forschungen* 4, 85f. (experiment 20; please delete experiment 19 because it is based on an error as explained in experiment 21). – Pohlmann, A. G. and F.W. Kranz. 1924. *Proceedings of the Society for Experimental Biology and Medicine* 21, 335.

[98] Wien, M. 1903. *Pflügers Archiv für die gesamte Psychologie* 97, 1. – Fletcher and Wegel. 1922. *Physical Review* 19(2), 533; and others (cf. *Jahresbericht der Physiologie* 1921/22, 374).

[99] McKenzie. 1922. *Proceedings of the National Academy of Sciences (USA)* 8, 188.

Weber's law has also been proved valid, at least for middle-range frequencies and intensities: A noticeable change of loudness requires a change of the stimulus intensity by a constant fraction.[100] At this point, like in stimulus *sequences* in general, another factor comes into play: the interim period. The succession of two sound units – like a succession of two brightnesses or *Tonigkeiten* – is a step: increasing, falling or continuing on the same level. The first sound sets a level which the other one stands out from. This level – physiologically a "silent track" – falls during the first few seconds of its existence. Thus, a step upwards, for example, increasing loudness, appears bigger if there is a longer interim period. But if we hear this phenomenon repeatedly, it becomes less manifest – the overlapping tracks counteract the falling of the level, even days later.[101]

The loudness of a sound is also influenced by the combination with other sounds, at least as much as by the antecedents. A partial filtered from a sound or chord always seems softer than an isolated tone of the same physical volume.[102] A part of the total energy – if there is nothing to "filter", the entire energy – is obviously used for the physiological overall process that becomes phenomenally apparent as sound (or chord) color.[103] It also seems as if a tone added to another tone, sound or chord immersed into it and a filtered partial stood out only partially from the masses of sound. How much we can filter – even with a maximum of effort – depends on the intensity and frequency and, to a great extent, on the listener. In general, lower tones absorb higher ones more easily than vice versa.[104] Consonance facilitates fusion. Both factors combined let the higher tone of an octave get easily lost in the overall sound.[105] It can be assumed that the homogeneity of sounds is also determined by certain distributions of intensity, which stabilize the overall structure – already physiologically – and make it difficult or impossible to filter partials. At equal objective conditions (and a maximum of attention), the loudness of a filtered tone varies significantly for different observers but seems constant for the same observer. However, it is not determined by the ability to analyze: People for whom the filtered tones always seem much softer than for others are often also able to filter still something under very unfavorable conditions.[106]

A chord as a whole also has a certain loudness which can be compared to the loudness of a single tone. If we alternate a tone and a chord on a piano, struck as constantly as possible, the chord seems fuller but hardly louder, at least not to the extent we would expect according to the increase in energy.[107]

Measuring experiments, however, are still to be conducted. Simultaneous pressure on several areas of the skin at objectively equal strength seems stronger

---

[100] Recently Guernsey. 1922. *American Journal of Psychology* 33, 554.

[101] Köhler, W. 1923. "Zur Theorie des Sukzessivvergleichs". *Psychologische Forschung* 4, 115–175.

[102] Stumpf, C.: *Tonpsychologie* 2, 418ff.

[103] Eberhardt, M. 1922. *Psychologische Forschungen* 2, 346–367.

[104] Stumpf, C.: *Tonpsychologie* 2, 227ff., 421; *Beiträge* 5, 141f. – Wegel, R. L. and C. E. Lane. 1924. *Physical Review* 23, 266–285.

[105] Helmholtz, H. v.: [6]*Tonempfindungen*, 104. – Stumpf, C.: *Tonpsychologie* 2, 352ff.

[106] Eberhardt, M. 1922. *Psychologische Forschungen* 2, 346–367.

[107] Stumpf, C.: *Tonpsychologie* 2, 423ff.

than on a single spot (the filtered pressure, however, seems weaker than the isolated one, like a tone in a chord).[108] It is not easy to ignore the differences in richness and expansion which are always implied. The ear may in fact behave differently than the pressure sense because the sound is not louder when we hear it binaurally instead of monaurally (see above). Assuming that a part of the energy in chords is used for "coupling", we can conclude from the measurements by Wegel and Lane[109] that two equally strong tones sounding together influence each other even more strongly the lower and the less different the frequencies are; the two-tone chord will be less superior to the chord regarding loudness, and more fused. This is in accordance with other experiences, at least regarding fusion.

Terminology

Inconsistent terminology, which still occurs in recent literature despite extensive debates[110], makes comprehension difficult and lets theoretical opposites appear more incompatible to outsiders than they are. In what has been discussed above, physical (and physiological) factors are separated from phenomenal ones as clearly as possible already by the terms used. That is why we, for example, only speak about the "frequency", but not about the "pitch" of a wave, or about the "loudness" instead of the "volume" of a phenomenon. The attempt has been made to keep the denomination of phenomena and their characteristics purely descriptive and free from theory. For the factor which becomes most apparent in musical sound and provides the basis for the similarity to an octave, "tone relatedness" and "consonance", the term "*Tonigkeit*" has been introduced because "pitch" (Köhler) refers to another aspect of the phenomena and "musical quality" of just "quality" (Stumpf, Révész and others) seemed to be too theoretical. (Tonality has been established in English although the German "*Tonalität*" in a musical sense is also essential in psychology). Phenomenally, characteristics which *seem* to be quantitatively gradable in some way – apart from their manifestation – can be distinguished from mere qualitative ones. The former comprise brightness, height, weight, density, loudness, distance (length of the step) and sound width; the latter include timbre, vocality, tonality, interval color and chord color. Among them, distance and interval color are successive characteristics, and sound width and chord color are simultaneous ones. (It also seems appropriate to distinguish them at first regarding their denominations.) If we also want to name the overall characteristic of the phenomenon, "character" would fit well: "sound character" ("noise, vocal, sound, tone character"), "interval character" and "chord character".

[108] Werner, H. 1922. "Intensitätspsychologie." *Zeitschrift für Psychologie* 10, 64.
[109] Wegel, R. L. and C. E. Lane: *Physical Review* 23, 266–285.
[110] In particular W. Köhler: "Akustische Untersuchungen III", 181ff. – Stumpf, *Beiträge* 8, 51.