Aspects of addressing headphone transfer characteristics by loudness comparisons

Florian Völk [1,2]

[1] *Bio-Inspired Information Processing, Institute of Medical Engineering, Technical University of Munich*

[2] *WindAcoustics UG, Windach, Germany; Email: voelk@windacoustics.com*

Introduction

The transfer characteristics of an audio playback system with respect to a reference scenario can be addressed perceptually as follows: narrowband sounds at center frequencies covering the frequency range of interest are presented alternately by the system under test and the reference. Listeners are then instructed to adjust the level of the test sounds so that both systems are perceived equally loud (Beranek 1949, Zwicker and Gässler 1952, Pritchett 1954). The resulting frequency-dependent correction levels have recently been termed loudness-transfer functions (LTFs; Völk et al. 2011, Völk and Fastl 2011).

Two prominent LTFs are the target filter gains for the (perceptual) free-field and diffuse-field equalization of headphones, where the reference scenarios are frontally-incident plane waves or a diffuse sound field, respectively (for example Fastl and Fleischer 1978, Theile 1986). An aspect likely relevant when measuring LTFs between headphones and another method of audio playback is the repeatedly-reported difference in auditory-canal sound-pressure level at equal loudness for headphone versus loudspeaker presentation (e. g. Munson and Wiener 1952, Fastl et al. 1985). As an example, figure 1 shows the results of Fastl et al. (1985): for low-frequency pure tones, two different headphones (open model: white squares, closed model: filled circles) elicit the same loudness as a loudspeaker in the free sound field if the sound-pressure level in the auditory canal is on average about 4 dB higher.

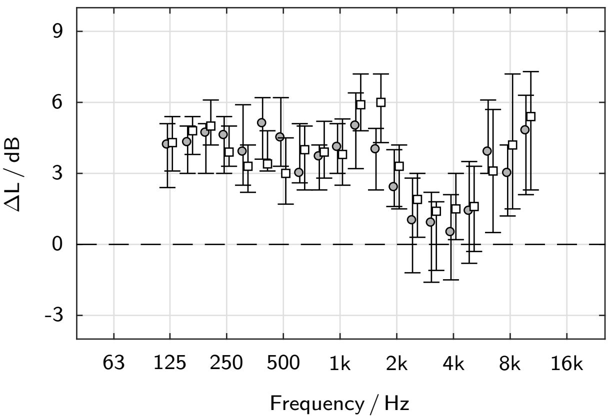

Figure 1: Auditory-canal level difference between binaural listening to diotic headphone and frontal loudspeaker presentation of pure tones at equal loudness (3.5 m distance to the loudspeaker, 70 dB SPL at the listening position, free sound field). Quartiles from Fastl et al. (1985): filled circles indicate "closed", white squares "open" headphones.

This contribution discusses the data of figure 1 in the light of LTFs measured with headphones versus a loudspeaker reference (Völk and Fastl 2011, Völk 2013). The results indicate that LTFs depend on the positions of the corresponding hearing sensations (as suggested by Theile 1982), especially when one of them is located inside the head (cf. Völk 2013). While this assumption not necessarily invalidates the procedure of loudness comparison, its implications must be taken into account, especially when applying LTFs, for example as the target filter-gains for headphone equalization, or when evaluating headphone-based listening experiments in general.

The Association Principle

Völk (2013) hypothesized, based on LTFs measured with headphone-based binaural synthesis versus a loudspeaker reference, that equal auditory-canal levels of narrow-band sounds occur at equal loudness when the hearing-sensation positions are similar or identical for both presentation methods. A similar conclusion may be drawn from the association principle put forward by Theile (1980, 1982). Theile may be basically interpreted in that, while forming hearing sensations, the auditory system conducts the localization process before evaluating loudness.

The above assumption appears reasonable also from an engineering point of view (cf. Völk 2013): In typical acoustic-communication scenarios, the signals of interest (sound radiated by an electroacoustic transducer, speech, music, or natural sounds; as some examples) can be considered as being partly distorted during the propagation to the listener's eardrums: certain acoustic properties of the propagation channels from the sources to the eardrums are superimposed on the signal carrying the information of primary interest. From that perspective, also the relevant time-varying transfer characteristics of the listener's body (especially head and outer ears) can be regarded as disturbing the actual signals of interests and, respectively, the encoded information. Typical engineering approaches of extracting information from potentially distorted signals at the receiver include the estimation of the channel characteristics and the separation of channel characteristics and signal of interest.

In the specific case of auditory-information processing discussed here, the estimation of the channel characteristics is basically the localization process. In the simple case of a free sound field, the direction-specific and partly distance-related changes of the head and body transfer characteristics can be represented by the so-called head-related transfer functions (HRTFs). According to the

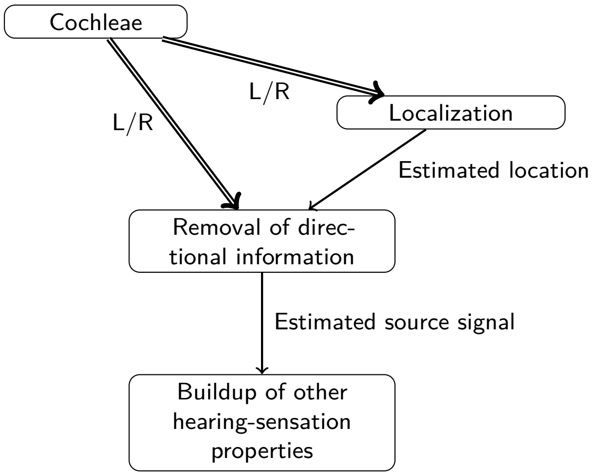

Figure 2: Descriptive and simplified interpretation of the information flow suggested by the association principle (Theile 1980, 1982), as interpreted by Völk (2013). Both cochlear "output patterns" (L/R) are analyzed for localization and modified before other hearing-sensation properties are built.

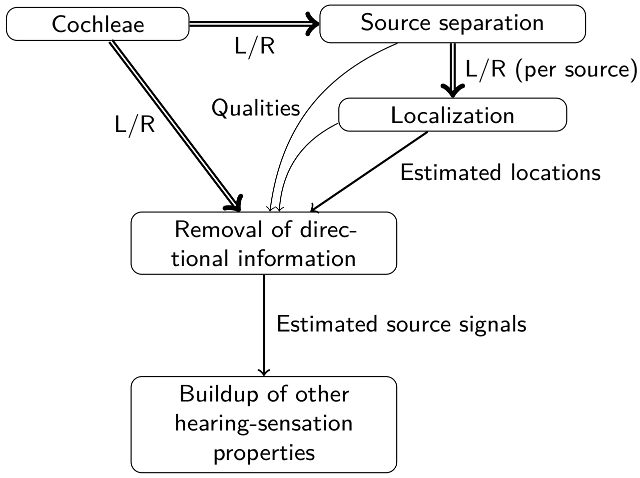

Figure 3: Working model of the hearing-sensation buildup as proposed by Völk (2013). In addition to the procedure described by figure 2, a continuous source separation and control paths propagating the estimated quality of the results of different model stages are introduced.

association principle, as well as in the above-mentioned engineering approach, the channel characteristics are to be separated from the signals of interest before the latter are being evaluated further. This somewhat descriptive and simplified interpretation of the association principle is illustrated as an information-flow chart in figure 2.

The human auditory system primarily detects sound-pressure variations at two typically time-varying positions, at the eardrums. These sound-pressure signals are first converted into mechanical vibrations in the middle ears and then encoded in electrical auditory-nerve potentials in the inner ears (in the cochleae; for a recent overview cf. Rudnicki et al. 2015). Figure 2 starts with the information processed by the cochleae, here referred to as cochlear "output patterns" (L/R), refraining from more closely specifying their actual physical and neurophysiological representation. This information is then processed and modified within the localization "stage", *before* being evaluated regarding other hearing sensations. Please note that the actual neurophysiological processes are more complicated and most likely not divided into separate stages per hearing-sensation property. However, this perspective appears reasonable and helpful for the discussion of information flow and information processing attempted here.

Schematic Model

From the assumptions described in the previous section, Völk (2013) derived a schematic working model of the hearing-sensation buildup, which is described and interpreted with regard to headphone vs. loudspeaker listening in this section. As a basis for this discussion, figure 3 shows the structure of the working model proposed by Völk (2013) and discussed in the following.

In order to be applicable to real-life scenarios, a localization model must initially separate contributions from different sound sources with potentially different positions in space. As typical scenarios vary over time, both, source separation and localization should take place continuously. Völk (2013) descriptively referred to these processes as a continuous source separation (upper right corner of figure 3) and a continuous spectro-temporal, binaural pattern matching, implementing the localization process (below the source separation in figure 3).

An important property of the model is that it includes the possibilities that source separation or pattern matching will not yield perfect results. To account for these situations, both stages are assumed to continuously estimate the quality of their results. The quality estimates are then taken into account in the following stage, in figure 3 referred to as "Removal of directional information". In short, Völk (2013) assumed that only the signal characteristics attributed to a source and identified as location information are subsequently removed and therefore not evaluated in forming other hearing sensations. This is implemented in the schematic model by the "Quality" parameters indicated by thin arrows in figure 3. Völk (2013) referred to the quality parameters as correlation coefficients, associated with corresponding basis functions.

Applying the above-described model structure to the auditory localization process has two major consequences:

1. The localization process may fail for a specific source, in which case the hearing sensation is assumed to be attributed a "fallback" location, presumably inside the head. However, there is no strict limit or threshold for a valid result; identified localization cues affect the hearing sensation position, even if not all cues are identified. It is assumed that in the latter case, lateralization or less-than-realistic externalization are observed (Völk et al. 2008, Völk 2009).

2. If a hearing-sensation position does not coincide with the corresponding source position(s), the unidentified localization information (channel characteristics)

is included in the signals evaluated regarding other hearing-sensation properties (as for example loudness), and will consequently affect them. In other words: the hearing sensations will be affected by the channel characteristics.

A question not discussed so far is how the "removal of channel characteristics" is actually realized. Essentially, it is possible to attenuate resonance peaks, to amplify the non-resonance regions, or to use a combination of the aforementioned methods, in order to flatten the magnitude-transfer characteristics. A discussion and possible motivations of potential implementations are given by Völk (2013). For the current study, the removal is assumed to be implemented as an attenuation of the transfer-function resonances, with an additional frequency-independent gain of 4.6 dB. This specific gain factor is selected because it is the average auditory-canal level difference between headphone and loudspeaker reproduction at equal loudness for 1 kHz tones reported by Fastl et al. (1985). The procedure may be interpreted in the sense discussed above as a combination approach, amplifying the non-resonance regions and slightly damping the auditory-canal resonance peak. As this procedure increases the level of correctly localized signal components vs. not localized (e. g. diffuse) contributions, which may be considered the signal-to-noise ratio, it appears to be a reasonable and beneficial approach for speech communication (cf. Völk 2013).

Applying the schematic model to the equalization of headphones by loudness comparisons, two different situations are to be compared: binaural listening to the reference loudspeaker and to the diotically-driven headphones under test. As known from everyday experience, these listening situations differ especially in the hearing-sensation positions: a single loudspeaker is typically heard at or close to its position, while diotic headphone playback causes a percept inside the head. According to the model, other hearing-sensation properties, here especially loudness, will not be affected by channel characteristics (head and outer-ear transfer functions) when listening to the loudspeaker. However, for diotic headphone playback, the hearing sensation occurs at the "fallback" position inside the head, for which the model predicts that the channel characteristics are not removed before building the other hearing-sensation properties. Therefore, no low-frequency amplification occurs. This leads to the global expectation of somewhat higher auditory-canal levels at equal loudness for headphone vs. loudspeaker listening. A more detailed evaluation of the schematic model is given by the computational plausibility check in the following section.

Test for Plausibility

The above-described working model was tested for its plausibility by simulating the loudness comparison between a frontal loudspeaker in anechoic conditions and headphones using a modified loudness-calculation procedure. As the procedure, the German national standard for calculating the loudness of stationary sounds (DIN 45 631 1991) was chosen (which may be considered a more detailed but fully compatible version of ISO 532 B 1975). This procedure (in its original form) takes into account what is called a "third-octave level correction according to the transfer characteristics of the ear", referred to as $A0$ (DIN 45 631 1991). In the procedure's operational mode for frontal sound incidence, $A0$ can be considered a rough approximation of the auditory-canal resonance and some high-frequency pinna effect. In terms of the above-discussed working model, the calculation procedure roughly reflects diotic headphone listening, where the localization defaults to the "fallback" location and the channel properties (outer-ear transfer characteristics) are not removed before building other hearing-sensation properties (as for example loudness). Here, this (original) implementation is referred to as the "headphone mode".

In order to mimic loudspeaker listening, the procedure of DIN 45 631 (1991) was modified: assuming the localization process succeeds and the channel characteristics are fully equalized before evaluating loudness by the procedure described above, $A0$ was removed from the calculation procedure by setting the correction level frequency-independently in each third-octave band to 4.6 dB. Keeping all other parameters of the "headphone mode", this is referred to as the "loudspeaker mode".

The actual test for plausibility was implemented as a model-based prediction of the auditory-canal-level difference at equal loudness reported by Fastl et al. (1985) and discussed in the introduction above. Therefore, in a first step, the loudness levels $L_{\mathrm{NGF,LS}}(f)$ of pure tones (1 s duration, 5 ms Gaussian gating) at the center frequencies of Fastl et al. (1985) and with $L_{\mathrm{LS}}(f) = 70\,\mathrm{dB}$ at the listening position were calculated in the "loudspeaker mode" (the results for both operational modes of DIN 45 631 and the sound synthesis were calculated using the respective functions of WindAcoustics Suite 2017). In the second step, the calculation procedure was repeatedly run again for each center frequency in the "headphone mode", with tones of varying level $L_{\mathrm{HP}}(f)$, until the calculated loudness levels for tones of the same frequency differed less than ± 0.1 phon between the operational modes, that is until $|L_{\mathrm{NGF,LS}}(f) - L_{\mathrm{NGF,HP}}(f)| < 0.1$ phon. This procedure resulted in a prediction of the frequency-dependent auditory-canal level difference $\Delta L(f) = L_{\mathrm{HP}}(f) - L_{\mathrm{LS}}(f)$ at equal loudness. Figure 4 shows the calculated results as a solid black contour, with the experimental results of Fastl et al. (1985) indicated by light gray medians and inter-quartile ranges in the background.

Comparing the calculated predictions (solid contour) with the experimental data of Fastl et al. (1985, gray symbols) reveals a structural similarity of both data sets, regardless of the headphone model (gray open squares or filled circles). This can be interpreted in that the model-based plausibility test cannot invalidate the schematic working model proposed by Völk (2013) and discussed in the previous section. However, the computational plausibility test does not prove the working model's validity, either. That being said, this study adds some material to the discussion about the auditory-canal level difference at equal loudness for headphone vs. loudspeaker listening, but was not able to eventually clarify the issue.

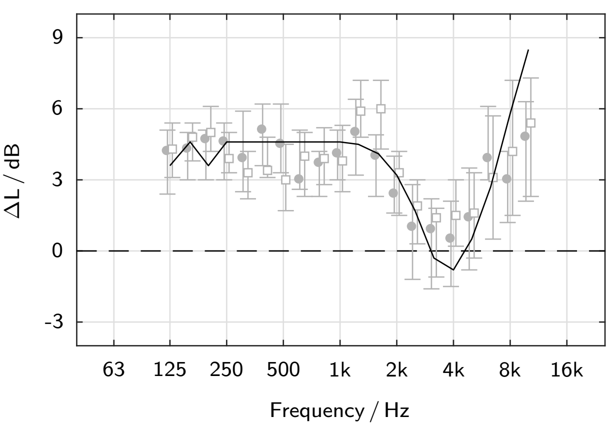

Figure 4: Auditory-canal level difference between binaural listening to diotic headphone and frontal loudspeaker presentation of pure tones at equal loudness (3.5 m distance to the loudspeaker, 70 dB SPL at the listening position, free sound field). Solid black contour: Predictions of the working model proposed here, implemented based on DIN 45 631 (1991). Light gray: Quartiles from Fastl et al. (1985): filled circles indicate "closed", white squares "open" headphones.

Summary and Conclusions

The present study addressed a potential explanation and a working model for the auditory-canal level difference at equal loudness in headphone vs. loudspeaker listening observed by different authors (e. g. Munson and Wiener 1952, Fastl et al. 1985). Explanation and model, originally proposed by Völk (2013), suggest, extending the association principle of Theile (1980), that loudness production and auditory localization are interrelated. The hypothesized impact of the perceived location on loudness becomes especially apparent when comparing headphone listening (with inside-the-head localization) and conventional loudspeaker listening (with typically clearly externalized hearing sensations).

The model's plausibility was confirmed in this study using a software implementation based on the partly-modified loudness-estimation procedure of DIN 45 631 (1991). However, since a conclusive final verification is still missing, the following conclusions and recommendations regarding the equalization of headphones by loudness comparisons can be given: headphones equalized by loudness comparisons do not generally provide the stimuli of the reference field; solely the average loudness of the reference field for narrowband sounds can be reproduced ("sound-field-equivalent levels do not ensure sound-field stimuli").

The full range of consequences and the effect's magnitude for arbitrary stimuli are not clear yet. Therefore, it is advisable to record the hearing-sensation positions when conducting listening experiments, and to keep the above considerations and dependencies in mind when interpreting the results of headphone-based studies.

References

Beranek L. L.: *Acoustic Measurements.* (John Wiley & Sons, Inc., New York, 1949)

DIN 45 631: *Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum. Verfahren nach E. Zwicker.* (Deutsche Norm, Beuth, Berlin, 1991)

Fastl H., H. Fleischer: Freifeldübertragungsmaße verschiedener elektrodynamischer und elektrostatischer Kopfhörer. Acustica **39**, 182–187 (1978)

Fastl H., W. Schmid, G. Theile, E. Zwicker: Schallpegel im Gehörgang für gleichlaute Schalle aus Kopfhörern oder Lautsprechern. In *Fortschritte der Akustik, DAGA '85*, 471–474 (DPG, Bad Honnef, 1985)

ISO 532 B: *Acoustics – Method for calculating loudness level.* (International Standard, International Organization for Standardization, ISO, Geneva, 1975)

Munson W. A., F. M. Wiener: In Search of the Missing 6 Db. J. Acoust. Soc. Am. **24**, 498–501 (1952)

Pritchett J.: The Subjective Free-Field Calibration of an External Telephone Receiver by the Equal-Loudness Method. Acustica **4**, 544–546 (1954)

Rudnicki M., O. Schoppe, M. Isik, F. Völk, W. Hemmert: Modeling auditory coding: from sound to spikes. Cell and Tissue Research **6** (2015)

Theile G.: *Über die Lokalisation im überlagerten Schallfeld*, PhD thesis, Technische Universität Berlin (1980)

Theile G.: Untersuchungen zur optimalen Kopfhörerentzerrung. In *Fortschritte der Akustik, DAGA '82*, 1247–1253 (DPG, Bad Honnef, 1982)

Theile G.: On the Standardization of the Frequency Response of High-Quality Studio Headphones. J. Audio Eng. Soc. **34**, 956–969 (1986)

Völk F.: Externalization in data-based Binaural Synthesis: Effects of Impulse Response Length. In *Int. Conference on Acoustics NAG/DAGA 2009*, 1075–1078 (Dt. Gesell. für Akustik e. V., Berlin, 2009)

Völk F.: *Interrelations of Virtual Acoustics and Hearing Research by the Example of Binaural Synthesis*, PhD thesis, Technische Universität München (2013)

Völk F., H. Fastl: Locating the Missing 6 dB by Loudness Calibration of Binaural Synthesis. In *131st AES Convention* (2011) (Convention Paper 8488)

Völk F., F. Heinemann, H. Fastl: Externalization in binaural synthesis: effects of recording environment and measurement procedure. In *Acoustics '08*, 6419–6424 (2008)

Völk F., A. Dunstmair, T. Riesenweber, H. Fastl: Bedingungen für gleichlaute Schalle aus Kopfhörern und Lautsprechern. In *Fortschritte der Akustik, DAGA 2011*, 825–826 (Dt. Gesell. für Akustik e. V., Berlin, 2011)

WindAcoustics Suite: *Modular software package for auditory-adapted signal and system analysis, real-time audio-signal processing, audio recording, and sound measurements.* (WindAcoustics UG [haftungsbeschränkt], www.windacoustics.com, 2017)

Zwicker E., G. Gässler: Die Eignung des dynamischen Kopfhörers zur Untersuchung frequenzmodulierter Töne. Acustica **3**, AB 134–139 (1952)