CHAPTER

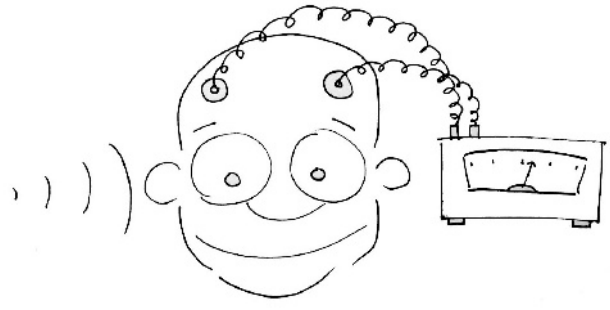
1

Introduction

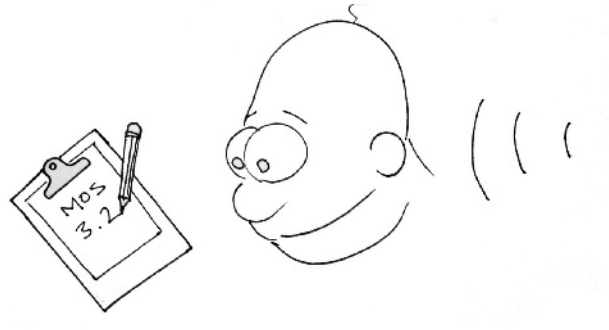
The aim of this book is to study the topic of perceptual evaluation of audio. Audio is very multidimensional in nature as is its perception. In order to study the nature of audio, it is possible to measure the physical characteristics of an audio signal in the acoustic or electrical domains. However, this characterisation of the physical audio signal does not tell us how the human auditory system will interpret and quantify it. In order to do this, a direct measurement of the human perception of the audio signal would be needed, as illustrated in Figure 1.1(a), but this is not yet possible.

An alternative means of assessing how listeners perceive an audio signal would be to ask them to quantify their experience. This is the most common form of perceptual evaluation that often takes the form of a formal listening test, as illustrated in Figure 1.1(b), and forms the core subject matter of this text. The manner in which listeners quantify their experience can occur on one of two levels, as defined

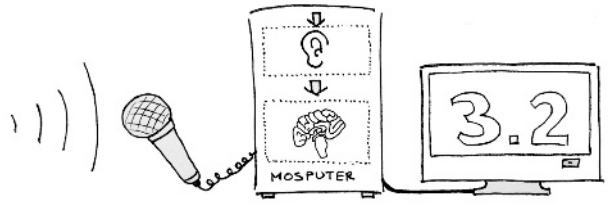
Perceptual Audio Evaluation – Theory, Method and Application Søren Bech and Nick Zacharov



(a)



(b)



(c)

Figure 1.1: Examples of different forms of perceptual evaluation. Reproduced by permission of Heikki Niemelä.

through the filter model to be introduced in detail in Section 4.1. The two measurement levels are defined as follows:

Perceptual measurement An objective quantification of the sensorial strength of individual auditory attributes of the perceived stimulus

Affective measurement An objective quantification of an overall impression of the perceived stimulus.

This text will address both of these measurement levels, though emphasis will be placed on perceptual measurement, as an objective means of evaluating the perception of an audio signal.

An alternative means of estimating the perceptual evaluation is through the use of predictive modelling techniques. This approach employs a perceptual model of the human auditory and cognitive system to predict the human response to an audio signal as illustrated in Figure 1.1(c). This approach is highly effective, as the aim of such predictive models often is to simulate the response of listeners participating in a listening test. Additionally, such models are very efficient compared to listening tests and can provide predictions of the perceptual evaluation in a very short time. As a result of time and cost efficiency, such methods are gaining popularity. However, as these methods are still under development, they cannot yet replace listening tests in many cases. A brief introduction to such perceptual evaluation methods, some of their applications and limitations is provided in Section 1.3.

1.1 Motivation for listening tests

Having established that one of the main means to study perceptual audio characteristics is through listening tests, it is now important to understand when such procedures should be applied. As will be illustrated during the course of this book, listening tests are arduous to perform and, as a result, are only undertaken when there is a clear need and benefit.

When considering the evaluation of audio and its qualities, the researcher should consider the different methods that are available. He should thus ask himself:

- Whether a measure of the physical signal will provide sufficient information?

- Does a direct measurement of the perceived audio quality exist?
- Can a suitable predictive model of perceived audio quality be identified for the assessment of the stimuli under consideration?

If it is established that the answer to all of these questions is no, then the researcher will need to consider performing a listening test, in which case he/she may read on.

In some cases, the experimenter will be working in a field that is well established and for which the norm is to perform a perceptual evaluation. Examples of such areas are audio and speech codec performance. In such cases, the researcher would benefit from the effort and expertise employed in the specification leading to a standardised approach for performing the listening test. The role and application of standards and recommendations will be discussed further in Section 1.2.

Having established that a listening test is required for perceptual evaluation, it is pertinent to have an idea of what to expect from the results. This is healthy for both the experimenter and his management to ensure that everyone knows what to expect.

Listening tests can provide some of the following pieces of information:

- Identify whether or not audio stimuli are perceptually identical.
- Establish whether a sample is perceptually equivalent, superior or inferior to another sample with regard to audio quality.
- Define to what degree a sample is superior to another in terms of audio quality.
- Establish which audio system is preferred.
- Establish whether an audio system is acceptable for a given task.
- Rate the performance of audio systems in a detailed manner employing a number of perceptual attributes.
- Define the absolute audio quality of an audio system.

Listening tests are not directly able to:

- Identify and locate the problem parameter of an audio algorithm.
- Identify which system will score highest in a Hi-Fi magazine evaluation.
- Identify what technical aspects of a competitor's audio system make it superior.

- Define how developers should improve their systems to obtain significant audio quality improvement.

However, with the use of advanced data mining techniques, it is possible for listening test data to be employed with other technical information to answer most of these latter issues. Such methods are, however, outside the scope of this book.

To guide the researcher through the process of identifying the correct approach to perceptual evaluation, the block diagram in Figure 3.1 (Chapter 3) is provided. The steps illustrated in this diagram form the basis for the chapters and sections of this book. Additionally, the mindmap provided in Figure 1.2 gives a detailed visualisation of the book structure to assist the reader in navigation.

Lastly, a short note regarding the scope of this book. This book is intended to cover audio and telecommunication applications and associated perceptual evaluation procedures. The focus has been placed on perceptual audio evaluation practices, although other methods are introduced in brief. The emphasis has been placed on listening-only procedures. As a result, topics such as conversational quality, discourse analysis, and so forth, are outside the scope of this text.

1.2 Role of standardisation

In the field of audio, a number of standards or recommendations exist that cover a wide range of topics from measurement devices through to perceptual evaluation methods for telecommunications or audio systems. A large number of standards pertain to perceptual evaluation of audio, either from the listening test perspective or relating to predictive models, which are discussed in the next section.

Typically, standards are developed when there is a large-scale need to address a problem within a field or industry. The need is usually driven by several parties when there is a perceived benefit from establishing a commonly agreed upon approach to addressing the problem.

Standards provide the benefit of an agreed upon approach, developed by experts in the field from both industry and academia. In terms of perceptual evaluation, this means that a methodology has been developed and verified as being applicable to the domain defined for that standard. This is of great benefit, as the experimenter does not have to develop his own evaluation method but can rely upon the

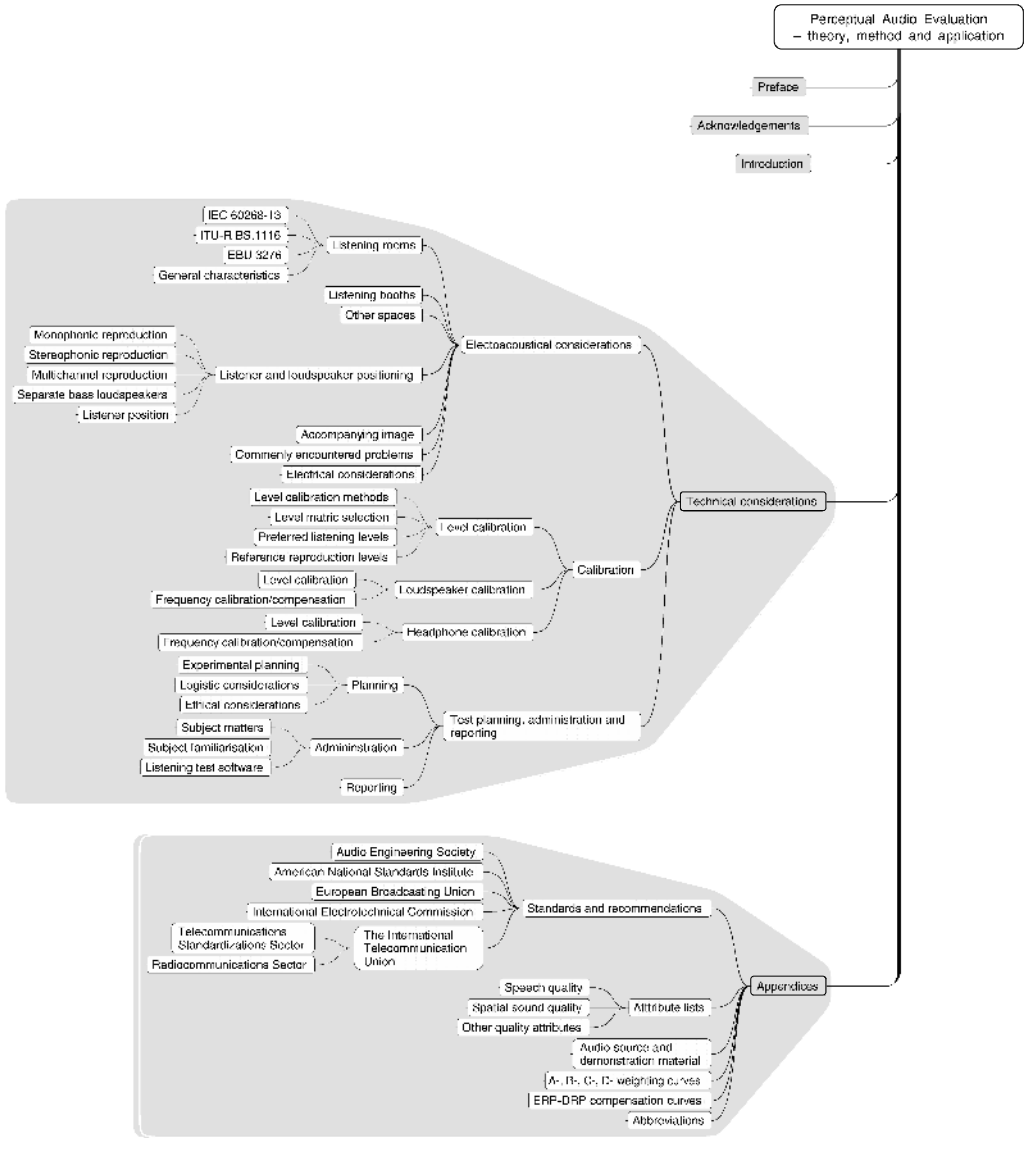


Figure 1.2: Mindmap illustration of the content and structure of this book.

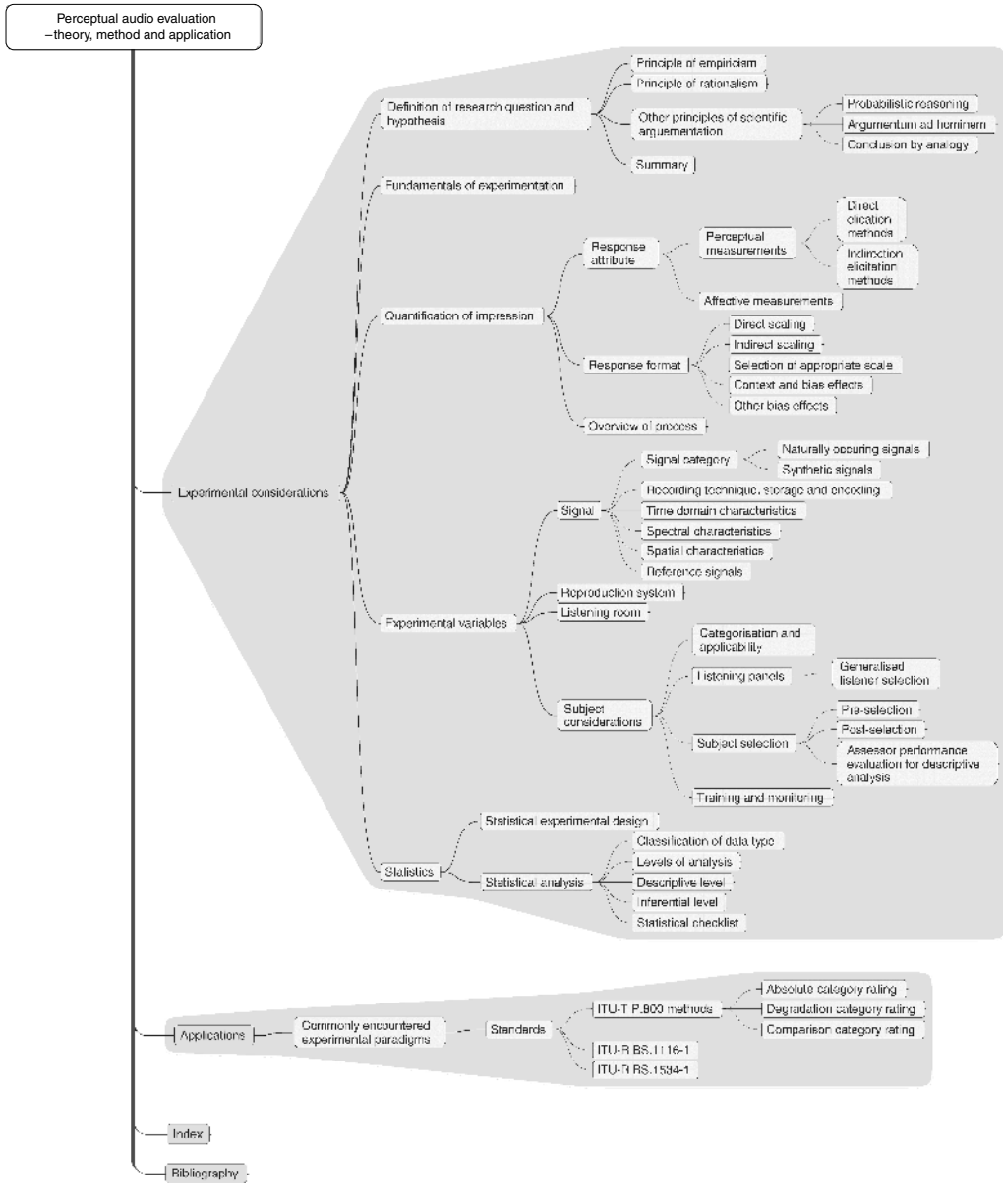


Figure 1.2: Continued.

expertise and prior analysis performed during the development of the standard. In such a standardised approach, a vital benefit is to provide an agreed upon method that allows for the comparison of results between different labs across the globe or at different times. This is a very valuable benefit, especially when large-scale experiments are undertaken and it simplifies the experimental planning to a large degree.

However, standards and recommendations require significant time for being developed and also require the consent of all parties involved. As a result, standardised methods are not always representative of the state-of-the-art methods in the field.

Additionally, not all matters are standardised, and this is due to several factors. Firstly, the process is very costly and time consuming. Additionally, as not all methods are commonly needed within an industry, only the key methods that require an agreement are studied and standardised.

As a result, in the field of perceptual evaluation, there are a number of key standards that define certain aspects of perceptual evaluation. These are quite well defined in terms of their domain of application and usage. When such a standard exists, the experimenter should consider whether the method is suitable for his task. If it is not, then other methodologies will need to be investigated. In either case, it is hoped that this book will guide the experimenter in establishing the best approach to be taken.

The experimenter should be wary of applying a standardised approach to an area outside the original scope of application of that standard. Additionally, modification of a standardised approach may be needed. In such cases, the experimenter should also be very wary and is advised not to state that the standardised approach has been followed, if this is clearly no longer the case.

While standards are not directly discussed in any particular section of this book, references to relevant standards and recommendations will be made throughout the text. Several standards organisations that provide valuable information of measurement methods, acoustic performance requirements or listening test methodologies will be referenced to. An example of some important standards from the International Telecommunications Union are presented in Figures 1.3 and 1.4. Appendix A is provided to summarise a number of key standards that relate directly to perceptual audio evaluation methods. This appendix aims to provide short guidance to the interested reader

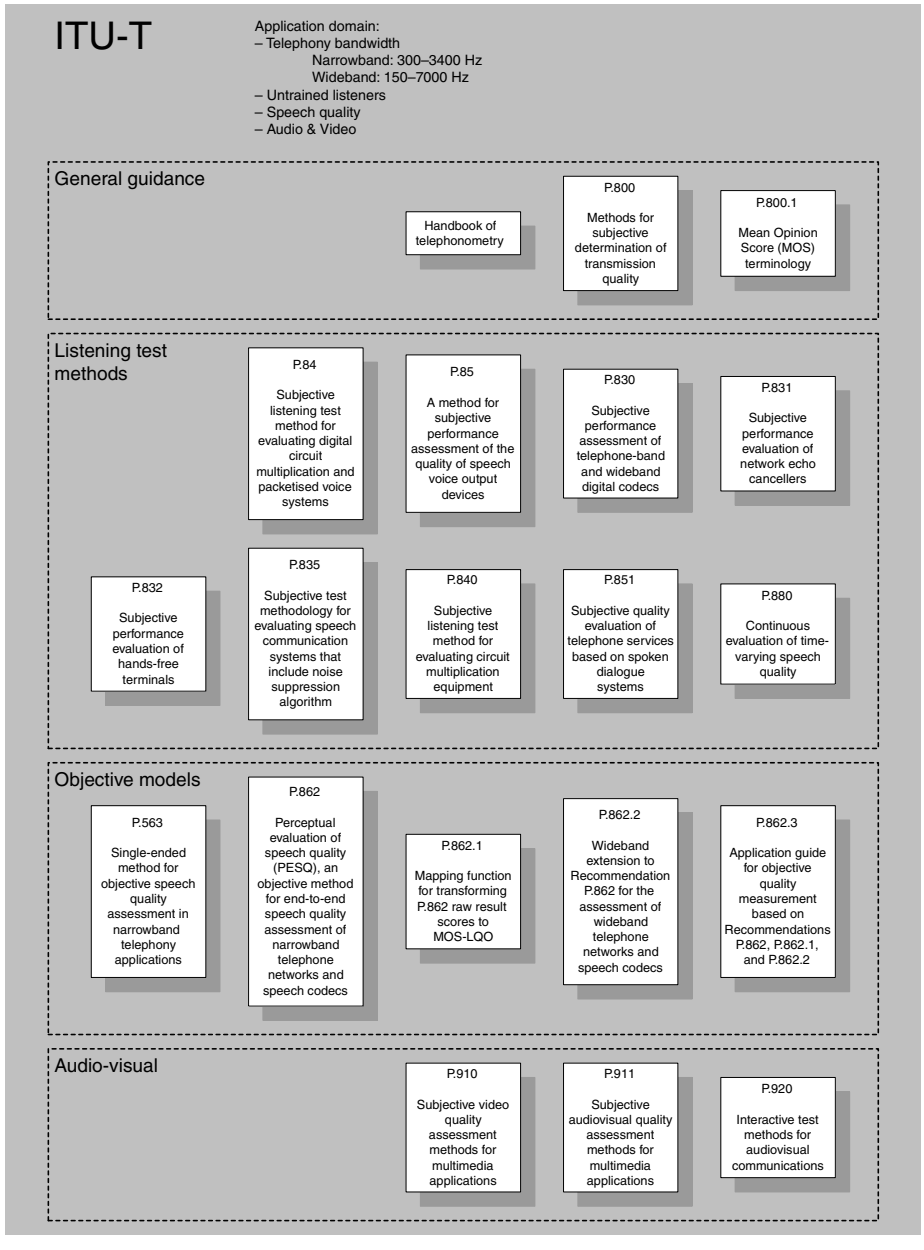


Figure 1.3: Summary overview of key ITU-T recommendations relating to perceptual audio evaluation.

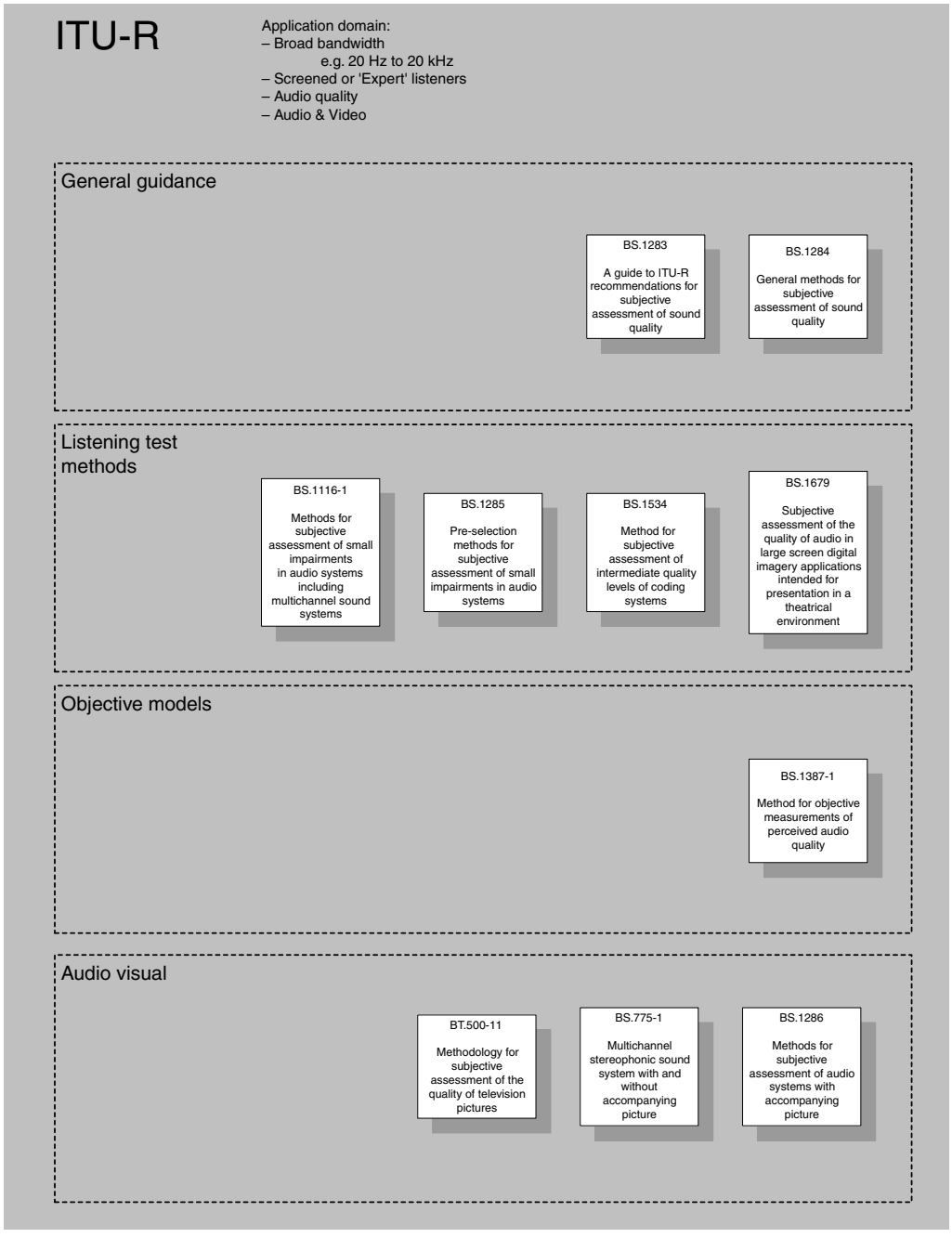


Figure 1.4: Summary overview of key ITU-R recommendations relating to perceptual audio evaluation.

regarding the scope of each standard or recommendation, prior to its acquisition.

1.3 Role of predictive models

One of the ultimate aims in the field of perceptual evaluation is to eliminate listening tests forever and replace them with predictive models that can accurately estimate human evaluation regarding the quality of audio signal. Development of such predictive models has been ongoing for many years. The present-day models typically comprise of a model of the human auditory system followed by a cognitive model to estimate the subjects scoring. Such models, nowadays always exist in the form of a computer program that analyses an input sound file, are very fast and easy to access and, as a result, are highly desirable.

Two categories of predictive models exist, namely, those that aim to predict a particular perceptual attribute, such as loudness, and those that aim to quantify overall performance, such as speech listening quality. In recent years, with the development of speech and audio codecs, there has been a desire to evaluate the performance of codecs and associated devices and this has led to the development and standardisation of numerous predictive models associated with speech and audio quality.

The benefits of such predictive models are quite clear and significant. As illustrated in this book, performing perceptual evaluation through listening tests is a time consuming and involved task, requiring significant knowledge of a number of different disciplines. The use of predictive models allows for the experimenter to obtain an estimate of the outcome of a listening test within a very short time, that is, real time to a few minutes typically and without the need to master a number of disciplines. This time and cost benefit is of great importance in addition to the near instantaneous access to such models, which require only rudimentary skill to operate.

The drawbacks of such models are, nonetheless, sufficiently important to heed. As with all tools, there are both correct and incorrect modes of usage. Predictive audio quality models normally have a particular domain of application beyond which their prediction accuracy is not known. Usage beyond the scope of application is risky and may provide misleading results, and thus the use of models in this manner is advised against. For example, models developed primarily for the

evaluation of narrowband speech have been trained extensively with different kinds of speech stimuli, codecs and other relevant stimuli. The applications of such models to audio codecs with music is not necessarily appropriate and may well lead to misleading prediction of the perceived audio quality.

Models that aim to study a unique attribute of human perception, such as loudness, can be found, for example, in [188, 305, 307, 473, 477]. The output of such models, for example, in the form of overall specific loudness, aim to predict the impression of a subject performing a scaling of stimulus loudness.

Several perceptual models have been developed for the prediction of overall audio quality characteristics. With regard to speech and telecommunication applications, they have been developed to predict the results of speech listening quality test as performed using an ITU-T recommendation P.800 absolute category rating (ACR) test paradigm. The perceptual evaluation of speech quality (PESQ) model has been developed for narrowband[1] telephony speech and provides high prediction accuracy in this application. This so-called intrusive model requires a reference signal. Its characteristics are described in ITU-T recommendation P.862 [236] and P.862.1 [242] as well as by its developers in Rix *et al*. [363] and Beerends *et al*. [41]. Recently, a wideband[2] telephony speech version of PESQ has been standardised in ITU-T recommendation P.862.3 [246].

The correct mode of application and the associated limitations are important to pay attention to and they are described in ITU-T recommendation P.862.2 [245] as well as in documents such as [289]. The experimenter who applies such models without taking heed of the application guides risks making prediction errors and drawing scientific conclusions that are ill-founded.

The so-called non-intrusive model for predicting the subjective quality of narrowband telephony applications requires no reference signal. Such models, as standardised in ITU-T recommendation P.563 [243], are of great interest in the monitoring of speech quality in live telephony network. However, owing to their non-intrusive nature, that is, lacking a reference signal, their performance and prediction accuracy are inferior to those of their intrusive counterparts, for example, PESQ.

[1] Narrowband telephony refers to a bandwidth of 300 Hz to 3.4 kHz.
[2] Wideband telephony refers to a bandwidth of 150 Hz to 7000 Hz [230].

For applications in low bit-rate audio codecs, the perceptual evaluation of audio quality (PEAQ) model has been standardised in ITU-R recommendation BS.1387 [213]. Details of this model and its applicability are discussed in [249, 416, 427].

As the reader can see, a number of models exist for predicting the results of listening tests in different and specific application domains. Often, within the scope of the target application, the prediction accuracy of such models is excellent. However, once beyond the scope of application, the prediction accuracy becomes unknown. Presently, there exists no unified predictive model that is able to cover all aspects of audio perception. Additionally, predictive models have not yet been developed for all aspects of audio perceptual evaluation. For example, spatial sound perception models are still evolving. As a result, it will still require some time before such predictive models can replace listening tests completely.

Until that time, listening tests will still be an essential part of perceptual audio evaluation, including the development and verification of such predictive models.