Ear and Hearing

The quest for ecological validity in hearing science: what it is, why it matters, and how to advance it

--Manuscript Draft--

Manuscript Number:	EANDH-D-20-00232
Full Title:	The quest for ecological validity in hearing science: what it is, why it matters, and how to advance it
Article Type:	Eriksholm Workshop: Ecological Validity
Section/Category:	Other
Keywords:	Ecological validity, Research, Hearing science, Hearing, Amplification, Laboratory study, Field study, Hybrid study, Outcome domains, Test variables,
Corresponding Author:	Gitte Keidser, PhD Oticon A/S Snekkersten, DK DENMARK
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Oticon A/S
Corresponding Author's Secondary Institution:	
First Author:	Gitte Keidser, PhD
First Author Secondary Information:	
Order of Authors:	Gitte Keidser, PhD
	Graham Naylor
	Douglas Brungart
	Andreas Caduff
	Jennifer Campos
	Simon Carlile
	Mark Carpenter
	Giso Grimm
	Volker Hohmann
	Inga Holube
	Stefan Launer
	Thomas Lunner
	Ravish Mehra
	Frances Rapport
	Malcolm Slaney
	Karolina Smeds
Order of Authors Secondary Information:	
Abstract:	Ecological validity is a relatively new concept in hearing science. It has been cited as relevant with increasing frequency in publications over the past 20 years, but without any formal conceptual basis or clear motive. The sixth Eriksholm Workshop was convened to develop a deeper understanding of the concept for the purpose of applying it in hearing research in a consistent and productive manner. Inspired by relevant debate within the field of psychology, and taking into account the World Health

	Organization's International Classification of Functioning, Disability, and Health framework, the attendees at the workshop reached a consensus on the following definition: "In hearing science, ecological validity refers to the degree to which research findings reflect real-life hearing-related function, activity, or participation." Four broad purposes for striving for greater ecological validity in hearing research were determined: A) better understanding the role of hearing in everyday life; B) supporting the development of improved procedures and interventions; C) facilitating improved methods for assessing and predicting ability to accomplish real-world tasks; and D) enabling more integrated and individualized care. Discussions considered the effects of variables and phenomena commonly present in hearing-related research on the level of ecological validity of outcomes, supported by examples from a few selected outcome domains and for different types of studies. Illustrated with examples, potential strategies were offered for promoting a high level of ecological validity in a study, and for how to evaluate the level of ecological validity of a study. Areas in particular that could benefit from more research to advance ecological validity in hearing science include: 1) understanding the processes of hearing and communication in everyday listening situations, and specifically the factors that make listening difficult in everyday situations; 2) developing new test paradigms that include more than one person (e.g. to encompass the interactive nature of everyday communication), and that are integrative of other factors that interact with hearing in real-life function; 3) integrating new and emerging technologies (e.g. Virtual Reality) with established test methods, and; 4) identifying the key variables and phenomena affecting the level of ecological validity in order to develop verifiable ways to increase ecological validity and derive a set of benchmarks to strive for.
Additional Information:	
Question	**Response**
"Public Access Policy" Funding Disclosure Please disclose below if you have received funding for research on which your article is based from any of the following organizations:	

Dear Brenda,

Please, find uploaded the consensus paper "The quest for ecological validity in hearing science; what it is, why it matters, and how to advance it" for the supplement on the 6th Eriksholm Workshop on "Ecologically valid assessments of hearing and hearing devices."

We hope you agree that the content of this and other papers of the supplement will make a great contribution to the field of hearing science. Any comments and suggestions you have for improving clarity and presentation in this paper would be most welcome.

Best regards,

Gitte Keidser & Graham Naylor

Sydney, May 29 2020.

The quest for ecological validity in hearing science: what it is, why it matters, and how to advance it

Gitte Keidser[1,2], Graham Naylor[3], Douglas S. Brungart[4], Andreas Caduff[5], Jennifer Campos[6], Simon Carlile[7,8], Mark G. Carpenter[9], Giso Grimm[10], Volker Hohmann[10], Inga Holube[11], Stefan Launer[12], Thomas Lunner[1], Ravish Mehra[13], Frances Rapport[14], Malcolm Slaney[15], Karolina Smeds[16]

[1] Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark

[2] Department of Behavioural Sciences and Learning, Linnaeus Centre HEAD, Linköping University, Linköping, Sweden

[3] Hearing Sciences - Scottish Section, School of Medicine, University of Nottingham, Glasgow, United Kingdom

[4] Walter Reed National Military Medical Center, Bethesda, USA

[5] Applied Physics Department and the Center for Electromagnetic Research and Characterization, The Hebrew University of Jerusalem, Jerusalem, Israel

[6] KITE - Toronto Rehabilitation Institute, University Health Network, Toronto, Canada

[7] School of Medical Sciences, University of Sydney, Sydney, Australia

[8] X-The Moonshot Factory, Mountain View, USA

[9] School of Kinesiology, University of British Columbia, Vancouver, Canada

[10] Auditory Signal Processing and Cluster of Excellence "Hearing4all", Department of Medical Physics and Acoustics, University of Oldenburg, Oldenburg, Germany

[11] Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, and Cluster of Excellence "Hearing4all", Oldenburg, Germany

[12] Department of Science and Technology, Sonova AG, Staefa, Switzerland

[13] Facebook Reality Labs Audio, Redmond, USA

[14] Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

[15] Machine Hearing Group, Google Research, Mountain View, USA

[16] ORCA Europe, WS Audiology, Stockholm, Sweden

All correspondence should be addressed to:

Gitte Keidser, Eriksholm Research Centre, Oticon A/S, Rørtangevej 20, DK-3070 Snekkersten, Denmark. gike@eriksholm.com, +45 48298900.

Financial disclosures/Conflict of interest:

This paper is the product of the 6th Eriksholm Workshop that was funded by the William Demant Foundation, including travel and accommodation for all authors.

ABSTRACT

Ecological validity is a relatively new concept in hearing science. It has been cited as relevant with increasing frequency in publications over the past 20 years, but without any formal conceptual basis or clear motive. The sixth Eriksholm Workshop was convened to develop a deeper understanding of the concept for the purpose of applying it in hearing research in a consistent and productive manner. Inspired by relevant debate within the field of psychology, and taking into account the World Health Organization's International Classification of Functioning, Disability, and Health framework, the attendees at the workshop reached a consensus on the following definition: *"In hearing science, ecological validity refers to the degree to which research findings reflect real-life hearing-related function, activity, or participation."* Four broad purposes for striving for greater ecological validity in hearing research were determined: A) better understanding the role of hearing in everyday life; B) supporting the development of improved procedures and interventions; C) facilitating improved methods for assessing and predicting ability to accomplish real-world tasks; and D) enabling more integrated and individualized care. Discussions considered the effects of variables and phenomena commonly present in hearing-related research on the level of ecological validity of outcomes, supported by examples from a few selected outcome domains and for different types of studies. Illustrated with examples, potential strategies were offered for promoting a high level of ecological validity in a study, and for how to evaluate the level of ecological validity of a study. Areas in particular that could benefit from more research to advance ecological validity in hearing science include: 1) understanding the processes of hearing and communication in everyday listening situations, and specifically the factors that make listening difficult in everyday situations; 2) developing new test paradigms that include more than one person (e.g. to encompass the interactive nature of everyday communication), and that are integrative of other factors that interact with hearing in real-life function; 3) integrating new and emerging

technologies (e.g. Virtual Reality) with established test methods, and; 4) identifying the key variables and phenomena affecting the level of ecological validity in order to develop verifiable ways to increase ecological validity and derive a set of benchmarks to strive for.

RATIONALE, SCOPE, AND APPROACH

Over the last few decades, hearing devices have evolved from straightforward amplifiers to highly sophisticated devices which respond to distinct environments to provide contextually relevant benefits to the wearer. Meanwhile, whilst many diagnostic and evaluation protocols have been computerized and automated, there has not been corresponding development in the procedures used for assessing a person's hearing ability and evaluating the benefit of increasingly complex hearing-related interventions. As early as 1988, the Working Group on Speech Understanding and Aging ("Speech understanding and aging. Working Group on Speech Understanding and Aging. Committee on Hearing, Bioacoustics, and Biomechanics, Commission on Behavioral and Social Sciences and Education, National Research Council," 1988) concluded that current audiometric tests were ineffective in determining a person's real-life hearing problems and benefit with hearing devices, pointing to the need to: 1) deliver test environments more comparable with the dynamic and reverberant real-world environments; 2) capture comprehension and; 3) consider cognitive processes involved in speech understanding. As shown in Figure 1, some studies concerning the use of more realistic environments or tasks in the research design were published prior to and during the 1990s. However, the last decade has seen a steep increase in such publications. The rise coincides with digital hearing devices becoming fully established and with the growing number of device features addressing varied, everyday demands. These device developments have prompted industry-based and academic researchers to, once more, point out the lack of evolution of the tests that continued to be used clinically and in research laboratories (Edwards, 2007; Jerger, 2009). The underlying problem was again suggested to be a lack of realism afforded by traditional test setups and tasks. Around the same time, Neuhoff (2004) advocated for an "ecological psychoacoustics", arguing that psychoacoustic investigations were traditionally limited to understanding the reaction of the auditory system to sounds, ignoring factors such as perception and cognition that would drive listening behaviors when detecting and recognizing sounds in the real world. Following this publication, a similar sharp increase in publications and conference presentations concerning ‘ecological validity’ of the research setting, stimuli, or outcome can be seen (cf. Figure 1).

Publications on hearing-related research concerning ecological validity have primarily reacted to calls for a more real-life approach by either 1) introducing novel and sophisticated test environments and tasks into laboratory studies, thus aiming to better replicate real-life listening situations (e.g. Grimm et al., 2016; Weller et al., 2016; Coene et al., 2018; Devesse et al., 2020), or 2) introducing more context-sensitive forms of measurement into field studies (e.g. Gatehouse, 1999; Wu et al., 2015; Wolters et al., 2016). While these studies have contributed greatly to moving the field forward in terms of making investigations more realistic in one way or another, they lack a shared conceptual basis, with the meaning of ecological validity and the purpose of striving for it in hearing-related research remaining unclear. For example, the term ‘ecological validity’ has been used to indicate both that the experimental context was more naturalistic (e.g. Zeni et al., 2019; Hadley et al., 2019) and that the approach had more face validity; i.e. provided the real-life information the researcher intended to obtain (e.g. Devesse et al., 2018; Decruy et al., 2019), and different motivations for adding more realism to research protocols have been floated.

To promote a more unified and streamlined research effort in the future, the aims of the sixth Eriksholm Workshop on Ecologically Valid Assessments of Hearing and Hearing Devices were, in broad terms, to:

- define the term ‘ecological validity’ as it applies to hearing-related research;
- outline purpose/s of striving for more ecological validity in this field;
- examine and discuss the variables and phenomena likely to affect the level of ecological validity in research studies assessing a person’s hearing ability (with/without hearing devices);

- summarize the current state of knowledge regarding the general level of ecological validity of various types of studies;
- identify knowledge gaps and research priorities.

This paper is structured as follows. The introductory sections motivate and present the consolidated definition of ecological validity and purposes of striving to improve it. Then we examine and discuss ecological validity in different types of studies (laboratory, field and hybrid), before presenting the state of the art in these domains as we perceive it, and considering how one might evaluate the level of ecological validity in a study (published or in planning). After a brief discussion of the effect a holistic approach to research designs may have on ecological validity, the final section lists knowledge gaps and future research priorities.

DEFINITION OF ECOLOGICAL VALIDITY

To draw meaningful conclusions from any research study, it is important to consider the validity of the study results. Most researchers are familiar with the concepts of internal and external validity; with the former examining whether a research study was designed, conducted, and analyzed appropriately to answer its research questions, and the latter examining whether findings of a research study can be generalized to other contexts. The concept of ecological validity is less familiar, but has in particular a long history in the field of psychology, where it is widely considered a concept that examines to what extent the results of a research study are related to, or predict, outcomes in situations occurring in everyday life. In those terms, it is generally thought of as a type of external validity. In psychology, the degree of ecological validity of a study is assumed to be closely tied to three methodological dimensions; the nature of a study's setting, types of stimuli implemented, and type of response used (Lewkowicz, 2001). According to Schmuckler (2001), the early debate on ecological validity centered, in particular, around the impact the experimental setting or environment had on the research (Brunswik, 1943;

Lewin, 1943). The debate resulted in the following classical, albeit narrow, definition of ecological validity: “ecological validity refers to the extent to which the environment experienced by the subjects in a scientific investigation has the properties it is supposed or assumed to have by the experimenter” (Bronfenbrenner, 1977). Brunswik (1943) also implied the importance of making the stimuli and response more realistic. The former point was echoed by Gibson (1960), and elaborated on by Neisser (1976), who stressed that real-life inputs typically consist of information that is temporally and spatially extended as well as multimodal. As pointed out by Schmuckler (2001), these three dimensions (setting, stimuli, and response) do not constitute an exhaustive list of factors involved in increasing the ecological validity of research but offer a potential starting point in a discussion and definition of ecological validity.

Hearing loss is a chronic health condition, and as such, is often contextualized within the World Health Organisation’s International Classification of Functioning, Disability, and Health (WHO-ICF) framework (WHO, 2001) (e.g. Kiessling et al., 2003; Lind et al., 2016; Iillum and Gradel, 2017; Lersilp et al., 2018; Manchaiah et al., 2019; Jaiswal et al., 2019). Similarly, hearing research findings, management strategies and outcome measures are increasingly interpreted within this framework (e.g. Psarros and Love, 2016; Ali et al., 2017; Convery et al., 2019; Alfakir et al., 2019). The WHO-ICF framework categorizes the function and disability of a person into three interrelated levels referring to the body (*structure or function*), the whole person (*activity*), and the whole person in a social context (*participation*). Workshop participants felt that ecological validity in hearing research could therefore usefully be related to the WHO-ICF framework. So, inspired by the definition of ecological validity offered by closely related psychological sciences, while integrating the WHO-ICF framework that is well established within the healthcare domain, we offer the following definition of the ecological validity concept when applied to hearing-related research:

"In hearing science, ecological validity refers to the degree to which research findings reflect real-life hearing-related function, activity, or participation."

It is worth noting that ecological validity is not a binary phenomenon that is either present or absent from a research study, but each study presents a certain level of ecological validity. As with the concepts of external and internal validity, the assessment of a study's level of ecological validity is ultimately based on subjective judgement, so the concept of ecological validity cannot be used to provide comprehensive objective criteria for experimental designs.

PURPOSES OF STRIVING FOR MORE ECOLOGICAL VALIDITY, AND POTENTIAL BENEFICIARIES

Studies to date that have aimed at achieving more ecologically valid findings seem to have been driven mainly by the notion that many of the established methodologies used to assess hearing and hearing devices lack sufficient realism to produce adequately meaningful findings about a person's hearing function, activity, or participation in real life. Such shortcomings in the state of the art could have implications for many stakeholders. Meanwhile, it is relatively rare to find explicit mention of why a lack of realism might be important. For these reasons, two sub-goals of the workshop were:

- To obtain a clearer picture of why striving for greater ecological validity of research findings is desirable (i.e. the purpose(s)), and
- To identify who will benefit from those efforts (i.e. the beneficiaries).

Overall, four different purposes of striving for greater ecological validity in hearing-related research emerged that participants agreed on. They are:

A. To better understand the role of hearing in everyday life.

B. To support the development of improved hearing-related procedures and interventions.

C. To facilitate improved methods for assessing and predicting the ability of people and systems to accomplish specific real-world hearing-related tasks.

D. To enable more integrated and individualized hearing healthcare.

Purpose A expresses a need for a better understanding of how people use their hearing when undertaking everyday activities in their environments, and how this is affected by impaired body function. In the case of impaired hearing, this means understanding the kinds of activity limitations impaired hearing causes and how they manifest, as well as the behavior of hearing-impaired people and their communication partners when challenged in everyday situations. Using the definition of fundamental, applied, and translational research suggested by Cooksey (2006), this need is in the domain of fundamental research.

Purpose B expresses a need for evaluation protocols that enable meaningful assessment of the individual's real-life hearing ability and benefit from hearing interventions (including devices, communication strategies, design of built environments, and other means). This will enhance the quality of evidence used to support development of and justification for more advanced hearing-related diagnosis, rehabilitation and screening procedures as well as interventions, and lead to improved hearing-related quality of life for those with unmet hearing needs. This need is in the domain of applied (Cooksey, 2006) research.

Purpose C expresses a need for more meaningful criteria to be established for the evaluation of, for example, eligibility for hearing-related benefit (e.g. subsidized intervention, insurance pay-outs), or ability to perform hearing-dependent tasks that are critical within certain professions (e.g. military, police force). Such criteria will help to maximize the ability of people with hearing impairment to participate in society. This need is in the domain of translational (Cooksey, 2006) research.

Finally, purpose D expresses a need for a more person-centered approach to understanding disability and needs, facilitating optimal intervention by considering the individual's general health, social connectedness, healthcare environment, and other overarching factors. This implies a broad view of the ecosystem and how it affects the individual's hearing health. This need is also in the domain of translational research, and concerns the removal of barriers to participation, but in terms of connected health and care systems beyond those solely related to hearing.

As for who will benefit from more ecologically valid assessments, the individuals and groups identified by workshop participants were consolidated into eight categories: 1) person with hearing needs (who may or may not have a hearing impairment); 2) people in immediate daily interaction with a person with hearing needs; 3) hearing-care professionals; 4) hearing researchers; 5) funders and policymakers; 6) product developers and marketers; 7) creators (e.g. film makers and designers of interactive games), and; 8) designers of the built environment. Table I lists the categories of beneficiaries identified and to what extent each would likely benefit from the pursuit of greater ecological validity by way of each of the four purposes outlined above. Two points of notice are that advancements targeting purposes A and B benefit all potential stakeholders, and that people with hearing needs and hearing-care professionals both stand to benefit from advancement targeting all four purposes. As exemplified by the papers in this issue, it is our hope that future publications in this area will clearly state the purposes for which their work advances knowledge and development, so that information and progress can more easily be consolidated and tracked, respectively.

At this point we would like to emphasize that it is not the purpose of this paper to suggest that all future studies should aim for a high level of ecological validity. There is nothing intrinsically superior about experiments that are more ecologically valid; rather, the methodological approach to a study should always be driven by the research objectives (see also Lewkowicz, 2001).

DEFINITIONS OF PRIMARY CONCEPTS

Some concepts are referred to repeatedly throughout this paper and across the papers making up this special issue. To promote comprehension and consistency, the most important concepts are collected and provided (in alphabetical order) with definitions in Table II. Terms in italics are themselves defined elsewhere in Table II.

TYPES OF STUDIES

It is customary to label research studies as either ‘laboratory’ or ‘field’ studies, and this distinction also inspired the structure of the workshop sessions. However, while workshop participants could agree that laboratory and field studies could be distinguished by their test environment (implemented vs. real-world) and the level of control of variables of primary interest (high vs. low), it proved impossible to reach agreement about what constituted a general border between the two types of studies. Instead it was decided to consider laboratory and field studies to fall, in their purest form, at opposite extremes on a continuum, with any study in between being referred to as a ‘hybrid’ study. Ecological validity in each of these three types of studies (laboratory, field, and hybrid) will be considered in the following sections of this paper.

For this purpose, the ‘model’ of a laboratory study assumed here is one in which:

- some situation in which human hearing function is believed to play a role is emulated in a laboratory
- participants are instructed to carry out some activity in the laboratory setup
- the activity encompasses a ‘task’, which may or may not be made explicit to the participant
- the experimenter controls at least all independent variables of primary importance in the experiment

- the experimenter is interested in one or more outcome domains, and implements outcome measures accordingly (e.g. speech recognition score, task completion time, eye gaze statistics)
- if relevant, assessment of an intervention is assumed to be made by comparing outcome with the intervention vs. without, or vs. an alternative intervention.

Whereas the ‘model’ of a field study assumed here is one in which:

- participants are instructed to carry out their normal daily activities in the usual manner, despite any additional burdens generated by participation in the study
- the only burdens imposed by the experimenter are for the purposes of monitoring activity or environment, or for eliciting participant impressions of situations and corresponding outcomes
- apart from selecting participants and setting design or intervention parameters (e.g. duration of field observation, hearing device operation), the experimenter controls no independent variables in the experiment
- the experimenter may attempt to monitor independent variables as they occur
- the experimenter is interested in one or more outcome domains and implements outcome measures accordingly (e.g. Ecological Momentary Assessment (EMA - Shiffman et al., 2008; Galvez et al., 2012), diaries, device usage statistics).

A hybrid study is a study in which the experimenter has control of some, but not all experimental variables. Commonly that would mean either low control of task, environment, or stimuli in a study conducted in the laboratory, or high control of task, environment, or stimuli in a study conducted in the field. In reality, pure forms of laboratory and field studies are rarely carried out; most studies possess some characteristics of a hybrid study.

ECOLOGICAL VALIDITY IN LABORATORY STUDIES

In this section we attempt to provide a relatively comprehensive overview of independent variables that can potentially be important to consider in relation to their effect on the level of ecological validity of a laboratory study as defined above.

In classical reductionist experiment design, the goal of the researcher is to evaluate the influence of systematic changes in one or more carefully controlled independent variables on the values of one or more carefully measured dependent variables. To the greatest extent possible, any variability that might result from variables other than the independent variables should be eliminated from the experimental environment. However, the cost of eliminating this variability is the risk that the particular combination of values for the independent variables selected for the experiment might not generalize to the relevant real-life scenarios that served as the underlying motivation for conducting the experiment.

When considering the effect of independent variables on ecological validity, it is important to distinguish between outcome domains (e.g. speech intelligibility, listening effort, affect, interactivity) and the outcome measures used to assess them (e.g. speech reception threshold, pupil dilation, self-report, behavioral synchrony). The level of control or variation in independent variables of the design may affect the phenomena elicited in the outcome domain of interest, and thereby the ecological validity of the experiment. But in addition, the method of measuring in the outcome domain of interest may itself affect the phenomena elicited. In the following, we examine each of these aspects separately.

How independent variables may affect ecological validity in laboratory studies

In Table III we consider how interactions between independent variables and outcome domains may influence the ecological validity of study outcomes. The list of independent variables, although long, is not exhaustive. It includes many common independent variables along with others identified during the workshop as being of some potential importance. Likewise, since there is no accepted categorization of outcome domains in hearing science that can be applied to the present exercise, we have chosen a few diverse outcome domains (speech recognition, listening effort, interactivity, and affective response) for the purpose of illustration. It is our hope that readers contemplating experimental design choices may find these examples helpful when determining which independent variables are most important for promoting ecological validity in their chosen outcome domain(s).

When grouping the independent variables into meaningful methodological dimensions, working groups at the workshop produced widely differing categorizations. Subsequently, we have attempted to rationalize and simplify these dimensions, specifically by referring to the three dimensions (setting, stimuli, and response) identified by Lewkowic (2001). However, these three were not found to adequately encompass the diversity of variables identified for hearing research, nor to be entirely helpful distinctions. For example, individual variables frequently mentioned during workshop discussions were not clearly represented in the three dimensions identified by Lewkowic (2001). As another example, workshop discussions revealed that ‘setting’ was too broad a term to adequately distinguish between variables related to the presentation of stimuli and the context of participation. The final dimensions of independent variables that participants reached a consensus on and their approximate mappings to those of Lewkowic (2001) are:

- Sources of stimuli [stimuli, setting]
- Environment (presentation of stimuli) [setting, stimuli]
- Context of participation [setting]
- Task [response]
- Individual [N/A].

For each of these five methodological dimensions, Table III lists in the first column the identified independent variables that are commonly used in hearing science. Where necessary, brief examples or comments are provided in the last column of the table to contextualize the variable. Note that ‘Age’ does not appear in the list of individual variables. This is because age as such is not a variable that directly influences outcomes. Variables typically mediated by age (e.g. cognitive abilities, sensory abilities) are included explicitly, which is a more parsimonious approach, but requires one to make correct associations between age and its likely consequences. Similarly, ‘Demographics’ is absent, as the most common components of this variable that are expected to directly affect the ecological validity of outcomes (e.g. education/occupation, cultural background, and disease burden) are explicitly included in the list.

For each independent variable and example outcome domain, we have in the table indicated our best guess as to the extent to which the variable affects the ecological validity of outcomes in that outcome domain. The symbols in the interacting cells indicate: X = very likely (based on research or logic), ? = might, but not enough research to state, and o = probably not (mostly based on logic not research).

Looking across the variables and their symbols listed for each dimension, it would appear that the ecological validity of outcomes in the domain of speech recognition would be particularly affected by variables related to the environment, whereas the ecological validity of outcomes in the domain of affective response would be particularly affected by variables related to sources of stimuli and context of participation. It would further seem that if a researcher wishes to manipulate and/or comment on the ecological validity of a study that aims to investigate listening effort, variables in all five methodological dimensions should be carefully considered. It can also be deduced from the table that there is still a lot to be learned about the potential effect of the context of participation on the ecological validity of outcomes in the domain of speech recognition, and of environment in the domain of affective response.

How the measurement of dependent variables may affect ecological validity in laboratory studies

Here we examine the second set of interactions, namely between the method of outcome measurement (i.e. dependent variable) and the level of ecological validity of the phenomena elicited in the corresponding outcome domain. Note that we are not concerned here with the quality of a measurement (reliability, precision, bias, etc.), since that is an issue of internal validity, not external or ecological validity. Rather we aim to illustrate how the method of measurement itself may affect the phenomena one wishes to observe.

It is customary to categorize outcome measures according to the modality they make use of, namely behavior, physiology, and self-report. It may be possible in principle to measure any outcome domain via any of these three modalities, therefore the observations made below are not related to any specific outcome domains.

- Behavioural measurement

Participant behavior intrinsic to the experimental task may be used as a direct source of data. Examples of this include:

- speaking in an unscripted conversation between two participants, where derived metrics (e.g. turn-taking statistics, dialogue repair events, voice stress) are used as the dependent variables,
- gesture and body movement, when the participant is not instructed to move, but not prevented from doing so, may be used in the same way as described above for speech,
- with an experimental task that is multi-faceted (e.g. in a virtual reality (VR) setup, cross the road while conversing with a partner), trade-offs between sub-tasks may be used.

In itself, this type of measurement poses no threat to ecological validity. However, it is possible that participants who are aware that their behavior is being monitored may alter their behavior, whereby ecological validity would be compromised.

Behavior extrinsic to the task may also be utilized as an indirect source of data. Examples include:

- speaking to report task response (e.g. repeating the sentence heard, judging sense/nonsense of heard phrase, reporting direction of heard sound),
- gesture and body movement to directly represent task response (e.g. pointing in the direction from which a voice was heard).

This type of measurement poses a substantial threat to ecological validity, as the participant is required to alternate between in-task (e.g. perception, reasoning, interrogation of environment) and out-of-task (reporting) cognition and behaviors.

Whether the measurement modality is intrinsic or extrinsic to the participant task, requirements for intrusive technical apparatus can affect behavior. If that behavior is integral to performing the task, ecological validity may be compromised. For example, body-worn apparatus that is weighty or constrictive may cause participants' movement to be constrained and unrepresentative. On the other hand, if (for example) movement analysis is carried out via analysis of video recordings, no such problems occur.

- Physiological measurement

Physiological measures offer an attractive route to capturing participant responses with a high level of ecological validity, as they are considered fundamental indicators of body state and function. However, in most cases substantial threats to ecological validity arise from the technical equipment needed to acquire the data. This is because intrusive equipment, such as mounted electrodes and wearable motion trackers, may well affect behavior, similarly to what was noted above for behavioral measurements. Likewise, in the well-known ‘white coat effect’, mere attendance at a physician’s office for a blood pressure test has the effect of raising one’s blood pressure (Parati & Mancia, 2003). Similarly, ‘biofeedback’ techniques demonstrate that (at least in the presence of feedback), a degree of volitional control is possible over physiological markers such as heart rate and blood pressure (Williamson & Blanchard, 1979). All these types of effects may in principle reduce the ecological validity of physiological measurement outcomes. Meanwhile, note that the often massive moment-to-moment variation in physiological variables that occurs during natural behavior is not in itself a threat to ecological validity, but if behavior or circumstances are artificially constrained in order to minimize the resulting physiological ‘noise’, then the ecological validity of the resulting experiment is again at risk.

- Self-report

Here we regard self-report (e.g. ‘how easy was it to follow the talker?’) as an additional task, extrinsic to the experimental task (e.g. ‘try to follow the talker’). Adding an extra task like this implies a high risk to ecological validity. The exception is retrospective self-report, for instance where a self-report is made after a block of trials or as part of a paired comparison paradigm, thereby effectively becoming a task of its own, distinct from the experimental task. However, even in such a design, the participant may, during execution of the task of interest, allocate resources to gathering impressions for use during the self-report task, or take on an artificial ‘self-observation’ behavior.

ECOLOGICAL VALIDITY IN FIELD STUDIES

Field studies of hearing and hearing devices are universally justified on the basis that their results will reflect the everyday lived experience of the participants, or in other words, that they provide a high level of ecological validity. While it is undoubtedly the case that field studies are ‘born with' much greater ecological validity than laboratory studies, the workshop participants were unanimous in the view that studies are not guaranteed to possess adequate ecological validity merely by virtue of being carried out in participants' everyday life, and that significant risks to ecological validity are present in many forms of field study. In this section we describe the sources and nature of such risks.

In contrast to the earlier discussion relating to laboratory studies, almost no threat to ecological validity arises from design aspects concerning the control of independent variables, because no attempt is made to control these variables, except insofar as participant selection represents experimenter control of 'Individual' factors as defined in Table III above. Thus, except for selection bias (see below), all the threats to ecological validity in field studies arise from the requirements of measuring dependent variables and monitoring independent variables.

Factors that may affect ecological validity in field studies

In the following we list and describe phenomena that can occur in field studies and may pose a threat to a study's ecological validity. These phenomena can be seen from two perspectives: one being distortions of behavior, the other being biased sampling of normal behavior patterns.

- Distortions of participant behaviour within everyday life

- Reactivity, which refers to a change on an outcome measure of direct interest to the study, caused by a participant's conscious engagement with the construct being measured. Note that this might also occur in laboratory studies, although it is less likely.
- Temporary avoidance of certain situations. This could be to avoid feeling socially awkward (e.g. staying home from a party due to carrying extra equipment or the need to divert attention away from the situation to execute self-reports), or to avoid feelings of non-compliance

(e.g. skipping a swimming lesson because it coincides with a time the participant is expected to execute a self-report).

- Deliberately seeking out certain situations that are normally never (or less frequently) experienced by the participant in question. This may be done in order to ‘test out’ the study apparatus, because these situations pose special problems for the participant, or because, having seen these situations listed in self-report outcome measures, the participant feels a duty to experience them.
- Temporary changes in behavior, within the participant’s everyday situations. This includes any behaviors that the participant believes are required in order to fulfil study tasks correctly.
- Sometimes the extra equipment (e.g. an external signal processing unit or an assistive listening device) and tasks carried by a participant elicit a positive form of curiosity from their acquaintances. This positive reinforcement may encourage abnormal behavior, for example conversations about the study, or demonstrations of the equipment by the participant. While such conversations may be very similar to those normally occurring (e.g. showing a friend a new gadget), it is nevertheless a conversation that would not have taken place in the absence of the study.
- In the case of studies in which the researcher is placed *in situ* within the participant’s everyday life (e.g. Wildemuth, 2017), the participant may exhibit abnormal (for that person) patterns of behavior and/or situations, due to a perceived need to conform to certain imagined norms.

● Biased sampling of everyday life

Sampling bias can take several forms:

- Undersampling of situations in which the participant is unable or unwilling to respond (e.g. situations of high cognitive or social load).

- Oversampling of situations in which responding is free of negative social consequences (e.g. self-initiated self-reports during quiet time alone). This may occur for example if participants perceive a requirement to achieve a certain number of responses per day.
- Oversampling of situations that participants judge to be 'interesting', based on their perception of the experimenter's aims.
- Oversampling of situations in which participants find it easier to comply with study tasks because it is easier to make judgements (e.g. when a contrast in hearing device settings has a clear effect on successful communication).

Selection bias is primarily driven by issues of who is willing to take part in and able to comply with such studies (willing to bear whatever burden is imposed, and able to behave and respond appropriately), i.e. a self-selection bias. This means the situations being sampled may not be representative of the intended target population, even if the ecological validity is high at the level of individual participants. Of course, selection bias is also present for laboratory studies, but probably operates to generate differently biased samples of participants.

ECOLOGICAL VALIDITY IN HYBRID STUDIES

As defined earlier, hybrid studies combine design features of pure laboratory and field studies. Hence a combination of the threats to ecological validity discussed in the previous two sections would apply to these studies. For illustrative purposes, in this section we present and discuss several study design features that are frequently found in hearing science literature, and that would be categorized as 'hybrid' according to the present scheme.

One design feature concerns directed behavior in the field, that is when a test participant is asked to do very specific tasks in their everyday environments and/or to perform tasks only in predefined everyday environments. An example would be a study in which participants are instructed to perform systematic comparisons of different devices or device modes during their normal daily activities. Although data is collected in participants' everyday environments, such a study does not qualify as a pure field study because the participant is required by the experimenter to consciously manipulate the study's independent variables. This is an issue that could only be overcome if the hearing device itself was programmed to enter different modes at different times, without informing the user. In this example the level of ecological validity of outcome measures can be affected both by distortions of participant behavior within their everyday life, and by interaction with independent variables in the methodological dimensions of 'task' and 'context of participation' (Table III).

Another common design feature of hybrid studies involves participants making retrospective reports of experiences from the real world while situated in a research milieu (lab, office, or clinic). Data may be collected via questionnaires, interviews or focus groups. Although the data typically refer to unrestricted behaviors in the field, they do not qualify as pure field data as participants are removed from the real-life environments at the time they report their experiences. The data are not pure laboratory data either as the experimenter did not have the necessary control of any independent variables that could be interacting with the experiences referred to. In such designs the level of ecological validity of outcome measures can be affected by biased sampling of everyday life situations and the outcome method measurement (self-report). In addition, retrospection biases can in this case negatively affect the ecological validity of outcome measures.

A final design feature that deserves a mention in this section, is the special case of making self-reports with reference to hypothetical situations. Many popular questionnaires, whether self-administered at home or administered by the experimenter in a research milieu, ask test participants to imagine how they would perform in hypothetical real-life situations. In this case, participants may during data collection be dislocated from the environment in question not only in a physical sense, but also mentally if the situation is unfamiliar to them. The extra level of ‘dislocation’ is a further threat to ecological validity of outcome measures in this case, beside the factors mentioned with the previous example.

ECOLOGICAL VALIDITY IN HEARING RESEARCH: STATE OF THE ART

In this section we aim to evaluate the level of ecological validity currently achieved in state-of-the-art test scenarios. Before going into detail, the question is considered from a high-level perspective, namely the WHO ICF categories of Body function, Activity, and Participation (WHO, 2001). We propose that laboratory studies conforming to the model described above are best suited to studying the Activity category (e.g. behavior). Such studies are not likely to be efficient for probing the ‘biological’ end of the Body function category (e.g. hair cell status and function), nor to provide meaningful insights about the ‘societal’ end of the Participation category (e.g. quality of life), although that might be a plausible future goal. Field studies of the type discussed here are also unlikely to be efficient for probing the ‘biological’ end of the Body function category, but they are probably the best vehicle for providing meaningful insights about the ‘societal’ end of the Participation category. With careful design, field studies may also be effective for illuminating aspects of Activity.

In the previous sections we have discussed various factors and phenomena that likely affect the level of ecological validity in laboratory, field, and hybrid studies. In the process, we have in Table III introduced a list of independent variables considered of potential interest across several outcome domains when conducting hearing research.

In this section we take a closer look at the same independent variables, with the aim of evaluating the level of ecological validity broadly achieved for each variable when implemented in pure laboratory and field studies. Hybrid studies are not considered here, as the independent variables in those studies are either controlled, as would occur in laboratory studies, or not, as would occur in field studies. While diagnostic testing as it is carried out in the clinic was not a focus of the workshop, we included it as a type of 'study' for the sake of comparison. This exercise provides a glimpse of the current state of the art in hearing research and clinical practice and helps to highlight future research priorities. By 'current state of the art', we mean the highest level of ecological validity that can be achieved with established equipment and procedures, even if this level is only achieved by a few research laboratories. The exercise is necessarily crude due to the numerous and varied outcome domains and measures in use, as well as subjective, since judging the level of ecological validity of a study is not an exact science.

Table IV repeats in the first column the independent variables presented in Table III; except here we have bundled the range of personality and demographic variables under the 'Individual' dimension into one row, as the achieved level of ecological validity was judged to be the same across these variables. In the second column, we have for each variable provided some examples of design features that are likely to support a high level of ecological validity of a study. In the final columns, we have indicated whether we judge the variable to have a 'low', 'medium' or 'high' level of representation of the real world in the best contemporary standard-of-care clinical, laboratory, and field settings.

Looking across the rows of Table IV, it can be seen that in agreement with our earlier hypothesis, field testing would seem to present a higher level of ecological validity than both laboratory and clinical testing, suggesting that if ecological validity has high priority in a study, the more time intensive and less controlled field tests currently offer the optimal form of data collection. However, if field studies are to achieve analytical power, they need to be equipped with greater abilities to monitor the values of their uncontrolled variables. Of particular note is that clinical testing is lagging behind in the 'Sources of stimuli', 'Environment' and 'Task' dimensions, whereas for laboratory testing there is some scope for improvement in the 'Context of participation' dimension. Variables in the 'Individual' dimension naturally have the highest level of ecological validity in clinical settings, where individual factors of the specific patient are automatically operating as they should. In research settings, there is a substantial risk that individual variables unaccounted for (or deliberately excluded) reduce the level of ecological validity.

The design features listed in the second column of Table IV suggest some potential strategies for supporting a high level of ecological validity, especially in laboratory studies. We note that many of these strategies are further detailed and discussed in the selection of papers found in this issue (e.g. Brungart et al., this issue pp. XXXX; Carlile and Keidser, this issue pp. XXXX; Grimm et al., this issue pp. XXXX; Hohmann et al., this issue pp. XXXX; Lunner et al., this issue pp. XXXX; Smeds et al., this issue pp. XXXX) and are still to be formally verified. It should also be noted that some variables are likely to affect the level of ecological validity of a study more than others. A valuable future exercise could thus be to use the information provided in the accompanying papers to prioritize both the methodological dimensions of variables, as well as the variables within each dimension, in terms of their importance when the goal is to achieve a high level of ecological validity of a study. It is likely that the outcome domain of interest will influence which variables are most important.

EVALUATING THE LEVEL OF ECOLOGICAL VALIDITY OF HEARING RESEARCH STUDIES

As noted earlier, ecological validity is not a relevant criterion by which to evaluate all studies, but for those studies where it is relevant for a given purpose (e.g. purpose A, B, C, or D introduced earlier), we attempt in this section to distil the products of the workshop into some practical recommendations.

As also noted earlier, ecological validity is not a binary concept, and even studies conducted entirely in the real world (field tests) are subject to phenomena that can threaten the ecological validity of research findings. In other words, researchers should be mindful that a study is not either ecologically valid or not, but that each study presents a certain level of ecological validity and that in reality it is probably impossible to carry out a research study that is free of all threats to ecological validity. At the moment there are no formal guidelines for how to determine the level of ecological validity a study presents, nor any set of objective criteria for how to quantify a study design as more or less ecologically valid. Defining such guidelines or criteria would foremost require some agreed understanding of what constitutes the ultimate benchmark for each test variable to maximize ecological validity of a study. As is evident from many of the papers in this issue, our knowledge of how far we can push variables and phenomena of interest to increase the ecological validity of a study is growing. But efforts are still in their infancy, and the knowledge and ideas exemplified in the papers in this special issue need to be formally consolidated. In the meantime, it is our hope that the thoughts presented in this paper (although necessarily subjective and descriptive) will be helpful to researchers in assessing the level of ecological validity in studies in which ecological validity is stated to be relevant. Furthermore, we encourage researchers, when publishing their own work of this type, to disclose if and how specific efforts were made to obtain a high level of ecological validity in their study. Each of the five methodological dimensions of independent variables (Sources of stimuli, Environment, Context of participation, Task, and Individual – cf. Table III) should be considered, as applicable. Authors should further disclose if and how specific efforts were made to reduce the effect of potential threats to ecological validity identified for the chosen modality of data collection (behavioral, physiological, or self-report), and, if applicable, disclose if and how distortion of participant behavior and biased sampling of everyday life in the field were managed.

As an example, Jensen et al. (2019) conducted an Ecological Momentary Assessment (EMA) study aiming to obtain information about the auditory reality of hearing aid users. In this study, experienced hearing aid users were equipped with an EMA system that was used to collect information about participants' experience with two hearing aid programs in the field. The EMA system was programmed to prompt participants to answer a set of questions every two hours, but participants could choose to delay or reject the task at such times, they could disengage the prompt alarm when they thought it was inappropriate to be interrupted, and they could also elect to answer the questions unprompted at any time. Because participants manipulated the hearing aid settings as they went about in their everyday environments, this study classifies as a hybrid study, and as such the level of ecological validity of the study is affected by both the interaction with independent variables listed in Table III and the phenomena related to field studies. For this study, the assessment of the level of ecological validity would be something like this: As no restrictions were imposed regarding the listening environments in which hearing aid settings could be evaluated, and participants could control when to do the task, the study overall had a high level of ecological validity in terms of ‘Sources of stimuli’, ‘Environment’, and ‘Context of participation’. The fact that participants consciously had to switch between hearing aid programs when in different listening environments reduces the level of ecological validity in the ‘Task’ dimension. Because participants were selected from established pools of presumably high-functioning and healthy test volunteers, the level of ecological validity in the ‘Individual’ category is considered low. The high risk to ecological validity naturally posed by the use of self-reports was partly managed by enabling participants to answer questions about the listening situation and experience in situ. However, the need to read questions and response options and to use a touch screen to perform the task, as well as the option of delaying the completion of the questionnaire discounted some of this gain. The objective measures collected of the environment posed no threat to ecological validity of the study. Finally, there appears to have been no attempt made to alleviate potential threats to ecological validity of the study caused by distortions of participant behavior within everyday life. On the other hand, sampling bias of normal behavior patterns was partly controlled by prompting the participant for responses with set intervals. On balance, taking the aim of the study into account, the ecological validity of the study is judged to be between medium and high. It should be emphasized that the assessment of ecological validity in any specific study is dependent on the study's aim as well as its experimental design.

ECOLOGICAL VALIDITY AND HOLISM

Holism refers to the treatment of a person as a whole, taking into account the individual's cognitive and physical well-being, their social network, and environment, and not just the symptoms of their disease. It is a concept that is beginning to make inroads into audiology (Bray, 2018), amidst growing appreciation of comorbidities between hearing loss and other chronic health conditions (Besser et al, 2008). Campos and Launer (this issue, pp. XXXX) present an in-depth discussion of holism in hearing healthcare. While the potential benefit of a holistic approach to treatment of any health condition seems self-evident, two questions in relation to integrating holism in hearing research were discussed during the workshop:

- Would it increase the ecological validity of a study?
- Would it influence how the variables outlined in previous sections affect the ecological validity of a study?

Following a discussion of reasons for and against embracing holism in hearing research, workshop participants agreed that the positives (e.g. give hearing health better context) were more likely to increase the face validity than the ecological validity of a study, and that the negatives (e.g. add complexity to the study design) could be minimized by organizing programs of research that address confounds through sequential and parallel studies. The general consensus was that integrating holism into a study design may or may not increase its ecological validity.

On the second question, the general consensus was that integrating holism into a study design does not pose any specific threat to how the variables and phenomena discussed in previous sections affect the ecological validity of a study.

KNOWLEDGE GAPS AND RESEARCH PRIORITIES

It is apparent from the results of this Eriksholm Workshop, that in those areas of hearing research that are concerned with how auditory abilities are put to use in everyday life, improvements in our understanding depend upon diverse and increasingly sophisticated methods of measurement (which includes methods of experimental and statistical control). In this section, we summarize key areas where knowledge is still limited, and list some priorities for research to move our field forward in terms of achieving more ecologically valid research findings. The list is by no means exhaustive, nor suggestive of order of importance. Further research recommendations are found in the accompanying papers of this special issue, where many of the research priorities listed here are also discussed in more depth.

Understanding the processes of hearing and communication in real life

Many variables and phenomena that are considered to affect the ecological validity of a study have been presented in preceding sections. They are selected based on the experience of researchers working in the field of hearing science for many years. However, it was agreed during the workshop that we possess incomplete understanding of the processes of hearing in real life, and of the factors that challenge people with hearing problems in everyday situations. We believe that a more refined conceptual understanding of these issues would be helpful for designing and evaluating studies where high ecological validity is relevant, as well as for developing new assessment tests for clinical applications. The development of such understanding would benefit from the use of qualitative methods involving hearing-impaired people and their families (Rapport & Hughes, this issue pp. XXXX). In particular, it was identified that further work is needed that goes beyond traditional qualitative techniques, to include novel methods that capture the transient ephemeral nature of listening. For example, real-time data capture and mobile methods (also known as 'walking interviews' - Kinney, 2017), allow for qualitative assessments that take place in situ.

Another area deserving of special attention is understanding the processes of hearing in interactive communication situations. It is generally believed that communication difficulty is the most disabling consequence of living with a hearing problem. As is evident from several publications in this special issue, assessments that provide information of high ecological validity about a person's communication ability are of particular interest in hearing science (e.g. Brungart et al., this issue pp. XXXX; Carlile & Keidser, this issue pp. XXXX; Grimm et al., this issue pp. XXXX; Lunner et al., this issue pp. XXXX;). Specifically, it is considered of high priority to steer away from traditional unidirectional test paradigms and develop new bidirectional assessment paradigms that encompass the interactive nature of everyday communications. Workshop participants agreed that to achieve this, apart from applying qualitative methods to refine our understanding of the processes of communication in real life, further work is in particular needed to understand: 1) the behaviors (e.g. body language, vocal effort, turn-taking) of interlocutors that lead to communication success, and how to effectively measure and characterize such behaviors in multi-person scenarios; 2) the acoustic qualities of sounds, beyond the signal-to-noise ratio, that challenge participation in everyday communication situations, and; 3) what imaging data (e.g. EEG, MEG, fNIRS) and other physiological measures may provide in terms of metrics of communication success and its underlying processes.

Unified and extended methodologies

Technological advances are continuously affecting how we collect data. As discussed in several papers in this special issue, new technologies are expected to enhance our ability to obtain more ecologically valid outcome measures, both in the laboratory and in the field (e.g. Caduff et al., this issue pp. XXXX; Mehra et al., this issue pp. XXXX; Slaney et al., this issue pp. XXXX). While the prospects are exciting, knowledge of how best to utilize these new technologies and integrate them with established test methods is lacking. This means that data of potential high importance are being collected using novel, unique and incompatible methodologies, making it difficult to consolidate findings across research groups.

One technology that was discussed during the workshop is the Ecological Momentary Assessment (EMA) approach, which is rapidly spreading as a means to collect data in the field (Holube et al., this issue pp. XXXX; Smeds et al., this issue pp. XXXX). EMA systems can capture a diverse range of objective and subjective data, and it was agreed that identifying and developing a core data set that could form a common minimal basis for future EMA data collection, where practical, should have high priority. Such a unified minimal set could consist of one or a combination of the following components: 1) objectively measured acoustic characteristics; 2) questions and response options, and; 3) temporal structure of assessment intervals.

Another accelerating technology that was discussed is the VR method (Hohmann et al., this issue pp. XXXX). Historically, VR techniques have been used to assess the impact that hearing impairment has on extremely complicated operational tasks, like engaging in combat in a tank or flying a helicopter. In recent years, VR technology has become more affordable and more capable of simulating unconstrained everyday environments. Amongst other potential advantages, VR is anticipated to make it possible to simulate acoustical, visual, and inertial components of everyday communication situations and multitasking demands with greater control than can be achieved in the field. This technological advance calls for parallel research to develop new task paradigms able to meaningfully reproduce the communicative and cognitive complexities of situations involving two or more people. Also, the inclusion of hearing devices into VR setups is not trivial, and more research is needed to establish criteria for, and methods to achieve, adequate acoustical veracity of sound fields and trade-offs against (for example) freedom of participant movement.

Strategies for increasing and evaluating ecological validity of studies

Strategies that are thought to support a higher level of ecological validity in hearing science studies are listed in Table IV above. At present, any proposed strategies for increasing ecological validity must be regarded as speculative, as there currently exists no evidence for: 1) what variables and phenomena are most important and reliable for supporting a high level of ecological validity; 2) whether this depends on the outcome domain of interest, and; 3) what constitutes the ultimate benchmark for each test variable to maximize the ecological validity. A high priority was identified for research addressing these questions and leading to the development of verifiable recommendations for how to increase the ecological validity of measures obtained in different outcome domains, and a set of benchmarks to strive for. In this context, a need was expressed for a tool or metric, which could be used to assess the level of ecological validity of a study. Presently, field studies are believed to be the best approach to obtain outcomes with a high level of ecological validity. However, there is a pressing need for examinations into how newer technology may be utilized to better monitor the values of uncontrolled variables in field studies, in order to provide them with sufficient analytical power that their results can inform concrete progress.

Ecological validity and holism

Holism is a relatively new concept in hearing science, and little is known about how integrating the concept into research designs affects the ecological validity of a study. It was recognized that in order to expand the knowledge base in this area, established methodologies used in hearing research would need to be advanced to enable a more integrative approach to measure hearing together with other health, social, and environmental factors, and that this would require knowledge from other and unfamiliar disciplines. For some examples, see Campos & Launer, this issue pp. XX; Carpenter & Campos, this issue pp. XXXX..

CONCLUSION

The sixth Eriksholm Workshop on applying the ecological validity concept in hearing science reached consensus on: 1) a definition of ecological validity; 2) four broad purposes of striving for ecological validity in hearing research and their beneficiaries; 3) the main variables and phenomena that threaten the ecological validity of research findings in laboratory, field, and hybrid studies; 4) strategies that, based on current knowledge, are expected to support a high level of ecological validity of a study, and; 5) a range of knowledge gaps that would benefit from future attention. It further developed some thoughts on how to evaluate the level of ecological validity of a study, and on the effect of integrating holism on the ecological validity of a study.

ACKNOWLEDGEMENTS

This consensus paper is based on intensive discussions on ecological validity in hearing research that took place at the 6th Eriksholm Workshop in August 2019. Gitte Keidser and Graham Naylor convened and facilitated the workshop and have contributed equally to the preparation of the consensus paper. All authors contributed equally to the discussions during the workshop and have reviewed and provided critical feedback to preliminary drafts of the paper. In addition, the following participants contributed substantially to the development of the sections on laboratory studies (Jennifer Campos), field studies (Inga Holube and Karolina Smeds), and state of the art (Douglas Brungart). The authors further thank the William Demant Foundation for funding the workshop that made it possible to meet in person and work together on developing the consensus views presented here. Contribution to this work was further supported by the Medical Research Council [grant number MR/S003576/1]; and the Chief Scientist Office of the Scottish Government (Graham Naylor); the USA Government (Douglas Brungart); and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 352015383 – SFB 1330 B1 (Giso Grimm and Volker Hohmann).

REFERENCES

Alfakir, R., van Leeuwen, L. M., Pronk, M., Kramer, S. E., & Zapala, D. A. (2019). Comparing the International Classification of Functioning, Disability, and Health Core Sets for Hearing Loss and Otorhinolaryngology/Audiology Intake Documentation at Mayo Clinic. *Ear and hearing*, *40*(4), 858–869. https://doi.org/10.1097/AUD.0000000000000662

Ali, A., Hickson, L., & Meyer, C. (2017). Audiological management of adults with hearing impairment in Malaysia. *International journal of audiology*, *56*(6), 408–416. https://doi.org/10.1080/14992027.2017.1305515

Besser, J., Stropahl, M., Urry, E., & Launer, S. (2008). Comorbidities of hearing loss and the implications of multimorbidity for audiological care. *Hearing research*, *369*, 3–14. https://doi.org/10.1016/j.heares.2018.06.008

Bray, V. (2018). A holistic approach to managing hearing loss and its comorbidities. *Hearing Journal*, *71*(11), 14,16,17. doi: 10.1097/01.HJ.0000549524.79131.ab

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *32*, 513–531. https://doi.org/10.1037/0003-066X.32.7.513

Brungart, D.S., Barrett, M.E, Cohen, J.I., Fodor, C., Yancey, C., & Gordon-Salant, S. Objective assessment of speech intelligibility in crowded public spaces. This issue, pp. XXX.

Brunswik, E. (1943). Organismic achievement and environmental probability. *The Psychological Review*, *50*, 255–272. https://doi.org/10.1037/h0060889

Caduff, A., Feldman, Y., Ishai, P.B., & Launer, S. Physiological monitoring and hearing loss: Towards a more integrated and ecologically validated health mapping. This issue, pp. XXX.

Campos, J.L., & Launer, S. From healthy hearing to healthy living: a holistic approach. This issue, pp. XXX.

Carlile, S., & Keidser, G. Conversational Interaction Is the Brain in Action: Implications for the Evaluation of Hearing and Hearing Interventions. This issue, pp. XXX.

Carpenter, M.G., & Campos, J.L. The effects of hearing loss on balance: A critical review. This issue, pp. XXX.

Coene, M., Krijger, S., van Knijff, E., Meeuws, M., De Ceulaer, G., & Govaerts, P. J. (2018). LiCoS: A New Linguistically Controlled Sentences Test to Assess Functional Hearing Performance. *Folia phoniatrica et logopaedica: official organ of the International Association of Logopedics and Phoniatrics (IALP), 70*(2), 90–99. https://doi.org/10.1159/000490050

Convery, E., Hickson, L., Meyer, C., & Keidser, G. (2019). Predictors of hearing loss self-management in older adults. *Disability and rehabilitation, 41*(17), 2026–2035. https://doi.org/10.1080/09638288.2018.1457091

Cooksey, D. (2006). *A review of UK health research funding.* The Stationery Office.

Decruy, L., Vanthornhout, J., & Francart, T. (2019). Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. *Journal of Neurophysiology, 122*(2), 601-615. https://doi.org/10.1152/jn.00687.2018

Devesse, A., van Wieringen, A., & Wouters, J. (2020). AVATAR Assesses Speech Understanding and Multitask Costs in Ecologically Relevant Listening Situations. *Ear and hearing, 41*(3), 521–531. https://doi.org/10.1097/AUD.0000000000000778

Devesse, A., Dudek, A., van Wieringen, A., & Wouters, J. (2018). Speech intelligibility of virtual humans. *International journal of audiology, 57*(12), 908–916. https://doi.org/10.1080/14992027.2018.1511922

Edwards B. (2007). The future of hearing aid technology. *Trends in amplification, 11*(1), 31–45. https://doi.org/10.1177/1084713806298004

Galvez, G., Turbin, M. B., Thielman, E. J., Istvan, J. A., Andrews, J. A., & Henry, J. A. (2012). Feasibility of ecological momentary assessment of hearing difficulties encountered by hearing aid users. *Ear and hearing*, *33*(4), 497–507. https://doi.org/10.1097/AUD.0b013e3182498c41

Gatehouse, S. (1999). Glasgow hearing aid benefit profile: derivation and validation of a Client-centered Outcome Measure for Hearing Aid Services. *Journal of the American Academy of Audiology*, *10,* 80-103.

Gibson, J.J. (1960). The concept of the stimulus in psychology. *American Psychologist, 15*, 694–703. https://doi.org/10.1037/h0047037

Grimm, G., Hendrikse, M., & Hohmann, V. Survey of self motion in the context of hearing and hearing device research. This issue, pp. XXX.

Grimm, G., Kollmeier, B., & Hohmann, V. (2016). Spatial Acoustic Scenarios in Multichannel Loudspeaker Systems for Hearing Aid Evaluation. *Journal of the American Academy of Audiology*, *27*(7), 557–566. https://doi.org/10.3766/jaaa.15095

Hadley, L.V., Brimijoin, W.O., Whitmer, W.M. (2019). Speech, movement, and gaze behaviours during dyadic conversation in noise. *Scientific Reports, 9*, Article 10451. https://doi.org/10.1038/s41598-019-46416-0

Hall, D.A., Smith, H., Hibbert, A., Colley, V., Haider, H.F., Horobin, A., Londero, A., Mazurek, B., Thacker, B., & Fackrell, K. (2018). Core Outcome Measures in Tinnitus (COMiT) initiative. The COMiT'ID study: Developing core outcome domains sets for clinical trials of sound-, psychology-, and pharmacology-based interventions for chronic subjective tinnitus in adults. *Trends in hearing*, *22*. https://doi.org/10.1177/2331216518814384

Hohmann, V., Paluch, R., Krueger, M., Meis, M., & Grimm, G. The Virtual Lab: Realization and application of virtual sound environments. This issue, pp. XXX.

Holube, I., von Gablenz, P., & Bitzer, J. Ecological Momentary Assessment (EMA) in audiology – current state, challenges, and future directions. This issue, pp. XXX.

Illum, N.O., & Gradel, K.O. (2017). Parents' Assessments of Disability in Their Children Using World Health Organization International Classification of Functioning, Disability and Health, Child and Youth Version Joined Body Functions and Activity Codes Related to Everyday Life. *Clinical medicine insights. Pediatrics*, *11*. https://doi.org/10.1177/1179556517715037

Jaiswal, A., Aldersey, H.M., Wittich, W., Mirza, M., & Finlayson, M. (2019). Using the ICF to Identify Contextual Factors That Influence Participation of Persons With Deafblindness. *Archives of physical medicine and rehabilitation*, *100*(12), 2324–2333. https://doi.org/10.1016/j.apmr.2019.03.010

Jensen, N.S., Hau, O., Lelic, D., Herrlin, P., Wolters, F., & Smeds, K. (2019). Evaluation of auditory reality and hearing aids using an ecological momentary assessment (EMA) approach. *Proceedings of the 23rd International Congress on Acoustics*, 6545-52. http://pub.dega-akustik.de/ICA2019/data/articles/000423.pdf

Jerger, J. (2009). Ecologically valid measures of hearing aid performance. *Starkey Audiology Series*, *1*(1), 1-4.

Kiessling, J., Pichora-Fuller, M. K., Gatehouse, S., Stephens, D., Arlinger, S., Chisolm, T., Davis, A. C., Erber, N. P., Hickson, L., Holmes, A., Rosenhall, U., & von Wedel, H. (2003). Candidature for and delivery of audiological services: special needs of older people. *International journal of audiology*, *42* Suppl 2, 2S92–2S101.

Kinney, P. (2017). Walking Interviews. *Social Research Update*, *67*, 1-4. http://sru.soc.surrey.ac.uk/SRU67.pdf

Lersilp, S., Putthinoi, S., & Lersilp, T. (2018). Facilitators and Barriers of Assistive Technology and Learning Environment for Children with Special Needs. *Occupational Therapy International*, *4*, Article 705946. https://doi.org/10.1155/2018/3705946

Lewin, K. (1943). Defining the 'field at a given time'. *The Psychological Review*, *40*, 292–310. https://doi.org/10.1037/h0062738

Lewkowicz, D.J. (2001). The Concept of Ecological Validity: What Are Its Limitations and Is It Bad to Be Invalid? *Infancy*, *2*(4), 437-450. DOI: 10.1207/S15327078IN0204_03

Lind, C., Meyer, C., & Young, J. (2016). Hearing and Cognitive Impairment and the Role of the International Classification of Functioning, Disability and Health as a Rehabilitation Framework. *Seminars in hearing*, *37*(3), 200–215. https://doi.org/10.1055/s-0036-1584410

Lunner, T., Alickovic, E., Graversen, C., Ng, E.H.N., Wendt, D., & Keidser, G. Three new outcome measures that tap into cognitive processes required for real-life communication. This issue, pp. XXX.

Manchaiah, V., Granberg, S., Grover, V., Saunders, G. H., & Ann Hall, D. (2019). Content validity and readability of patient-reported questionnaire instruments of hearing disability. *International journal of audiology*, *58*(9), 565–575. https://doi.org/10.1080/14992027.2019.1602738

Mehra, R., Brimijoin, O., Robinson, P., & Lunner, T. Potential of Augmented Reality Platforms to Improve Individual Hearing Aids. This issue, pp. XXX.

Neisser, U. (1976). *Cognition and reality*. Freeman.

Neuhoff, J.G. (2004) Ecological psychoacoustics: Introduction and History. In J. G. Neuhoff (Ed.) *Ecological Psychoacoustics* (pp 1-13). Elsevier Academic Press.

Parati, G., & Mancia, G. (2003). White coat effect: semantics, assessment and pathophysiological implications. *Journal of hypertension, 21*(3), 481–486. https://doi.org/10.1097/01.hjh.0000052458.40108.ad

Psarros, C., & Love, S. (2016). The Role of the World Health Organization's International Classification of Functioning, Health and Disability in Models of Infant Cochlear Implant Management. *Seminars in hearing, 37*(3), 272–290. https://doi.org/10.1055/s-0036-1584414

Rapport, F., & Hughes, S. Frameworks for change in hearing research: valuing qualitative methods in the real world. This issue, pp XXX.

Schmuckler, M.A. (2001). What is Ecological Validity? A Dimensional Analysis. *Infancy 2*(4), 419-436. https://doi.org/10.1207/S15327078IN0204_02

Shiffman, S., Stone, A.A., & Hufford, M.R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology 4*, 1-32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Slaney, M., Lyon, R.F., Garcia, R., Kemler, B., Gnegy, C., Wilson, K., Kanevsky, D., Savla, S., & Cerf, V. Auditory Measures for the Next Billion Users. This issue, pp. XXX.

Smeds, K., Gotowiec, S., Wolters, F., Herrlin, P., Larsson, J., & Dahlquist, M. Selecting scenarios for hearing-related laboratory testing. This issue, pp. XXX.

Speech understanding and aging. Working Group on Speech Understanding and Aging. Committee on Hearing, Bioacoustics, and Biomechanics, Commission on Behavioral and Social Sciences and Education, National Research Council. (1988). *The Journal of the Acoustical Society of America, 83*(3), 859–895.

Weller, T., Best, V., Buchholz, J. M., & Young, T. (2016). A Method for Assessing Auditory Spatial Analysis in Reverberant Multitalker Environments. *Journal of the American Academy of Audiology, 27*(7), 601–611. https://doi.org/10.3766/jaaa.15109

Wildemuth, B.M (Ed.). (2017). *Applications of social research methods to questions in information and library science* (pp. 209-18). Libraries Unlimited.

Williamson, D. A., & Blanchard, E. B. (1979). Heart rate and blood pressure biofeedback: I. A review of the recent experimental literature. *Biofeedback and self-regulation, 4*(1), 1–34. https://doi.org/10.1007/BF00998947

Wolters, F., Smeds, K., Schmidt, E., Christensen, E. K., & Norup, C. (2016). Common Sound Scenarios: A Context-Driven Categorization of Everyday Sound Environments for Application in Hearing-Device Research. *Journal of the American Academy of Audiology*, *27*(7), 527–540. https://doi.org/10.3766/jaaa.15105

World Health Organisation. (2001). *International Classification of Functioning, Disability and Health (ICF)*. Geneva: WHO. https://www.who.int/classifications/icf/en/

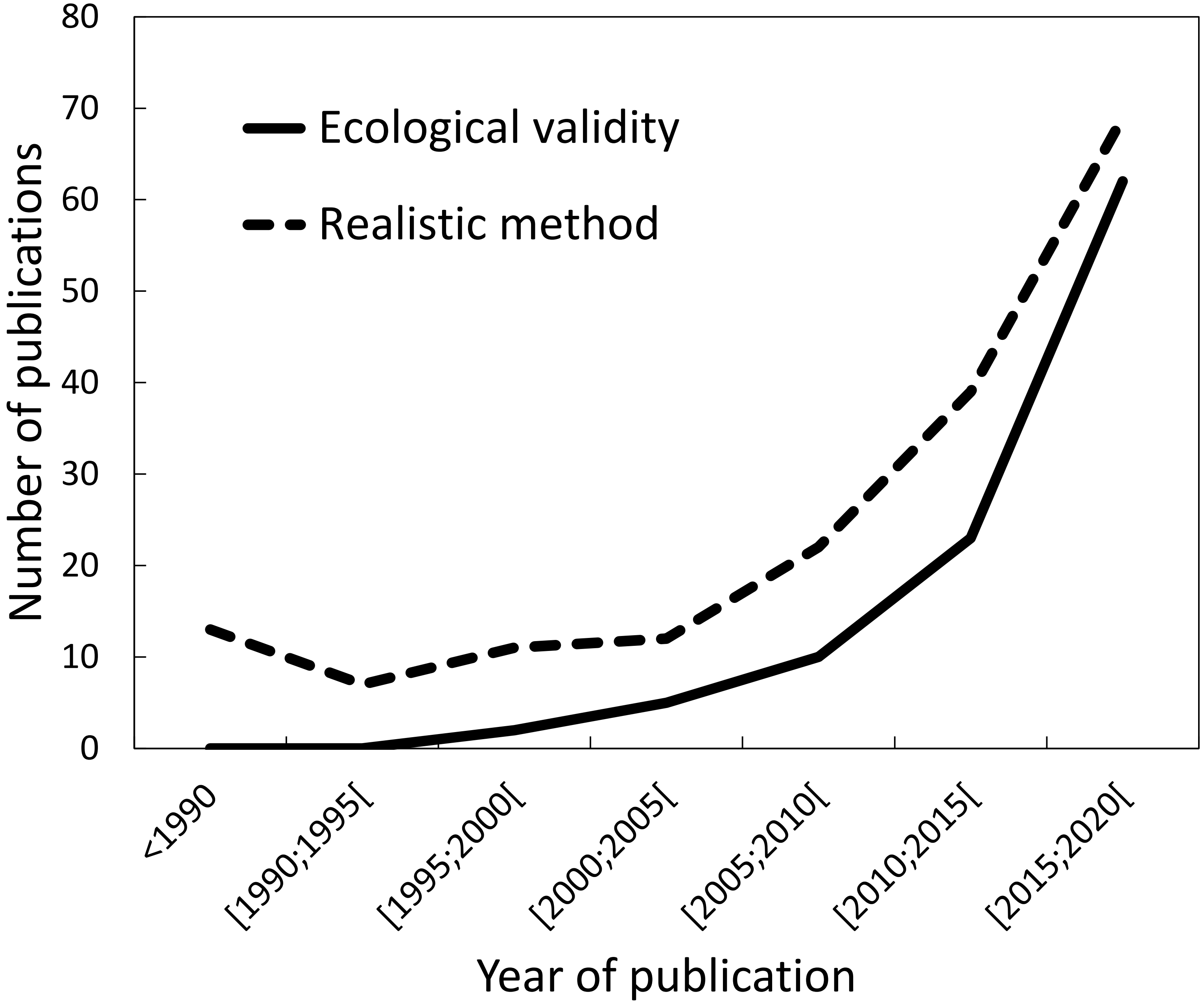
Wu, Y. H., Stangl, E., Zhang, X., & Bentler, R. A. (2015). Construct Validity of the Ecological Momentary Assessment in Audiology Research. *Journal of the American Academy of Audiology, 26*(10), 872–884. https://doi.org/10.3766/jaaa.15034

Zeni, S., Laudanna, I., Baruffaldi, F., Heimler, B., Melcher, D., & Pavani, F. (2020). Increased overt attention to objects in early deaf adults: An eye-tracking study of complex naturalistic scenes. *Cognition, 194,* Article 104061. https://doi.org/10.1016/j.cognition.2019.104061

FIGURE LEGENDS

Figure 1: Number of publications found on PubMed when using the combined search terms [(ecological OR ecologically) AND (valid OR validity) AND (hearing OR audiology)] - full line -, and [(realistic) AND (environment OR task OR method) AND (hearing OR audiology)] - bronken line - along a timeline showing five-year intervals post 1990.

Figure



80
70
60
50
40
30
20
10
0
Number of publications
Ecological validity
Realistic method
<1990
[1990;1995[
[1995;2000[
[2000;2005[
[2005;2010[
[2010;2015[
[2015;2020[
Year of publication

Table I: The likely beneficiaries of advancements in achieving more ecologically valid outcomes targeting each of the four purposes (A, B, C, and D).

Beneficiary	**Purpose A**	**Purpose B**	**Purpose C**	**Purpose D**
Person with hearing need	Y	Y	Y	Y
People in immediate daily interaction with a person with hearing need	Y	Y		Y
Hearing-care professionals	Y	Y	Y	Y
Hearing researchers	Y	Y	Y	
Funders and policymakers	Y	Y	Y	
Product developers and marketers	Y	Y		
Creators	Y	Y		
Designers of built environments	Y	Y		

Table II: The definitions of primary concepts used frequently in this special issue.

Term	**Description**
Everyday life	For a given individual, the subset of all *real-life* situations that are experienced with some significant frequency or have some significant importance.
Field study	A study in which the principal data-collection environment is each individual participant's *everyday life*, and in which primary stimuli and tasks are not controlled by an experimenter. See further description on pp. XXX
Hybrid study	A study in which the experimenter possesses control over one or more (but not all) of either the environment, the stimuli or the participants' *task*.
Laboratory study	A study in which participants are removed from their *everyday-life* environment and placed in an artificial one for the purpose of exposing them to controlled environments, *stimuli* and/or *tasks*, and obtaining pre-determined *outcome measures*. See further description on pp. XXX
Outcome domain	Any distinct aspect of function that could be assessed to determine whether an intervention has worked. (Hall et al., 2018)
Outcome measure	A measure, intended to reflect variation in a specified *outcome domain*, obtained through a replicable measurement procedure.
Participant task	A goal that a study participant might try to achieve. The task may be explicitly instructed by the experimenter or assumed by the participant.

Real-life or Real-world	Situations that are not controlled by an experimenter.

Table III: A list of commonly used independent variables in hearing science, grouped into the methodological dimensions of Sources of stimuli, Environment, Context of participation, Task, and Individual. For four example outcome domains, the independent variables are rated to show their suggested effect on the ecological validity of measures obtained in that domain; X = very likely, ? = might, o = probably not.

Independent variables	**Example outcome domains**				**Notes**
	Speech recognition	**Listening effort**	**Interactivity**	**Affective response**	
Sources of stimuli					
Characteristics of stimulus sources	X	X	X	X	e.g speech/other, diversity, familiarity, continuous vs. events
Characteristics of stimulus materials	X	X	X	X	e.g. monotonous, dynamic, neutral, emotional
For multimodal stimuli,	?	?	X	X	e.g. audio, visual, tactile

which modalities are subjected to controlled manipulation					
How other people are represented	?	X	X	X	disembodied voice -> real people axis. Includes potential for 'uncanny valley' effects
Environment (presentation of stimuli)					
Acoustic field	X	X	?	?	e.g. levels, SNRs, spatial fidelity, size of eventual sweet spot
Interaction of environment and hearing devices	X	X	o	?	degree to which the reproduced field (sound or other signal modalities) provokes the same device behavior as the real field would
Incorporation of	X	X	X	?	e.g. movement of sources

dynamic aspects					
Modalities included	X	X	X	X	e.g. visual, inertial
Context of participation					
Participant preparation	X	X	X	X	e.g. instructions, explanation provided for the purpose of the experiment, familiarization/training sequence
Semantic associations of the situation being simulated for the participant	X	X	X	X	e.g. does the participant ever take part in such a situation, does the participant have negative associations with it ("I always fail")
Motivation to take part	?	X	X	X	e.g. incentive, reimbursement, mode of recruitment
Familiarity with the lab	?	X	X	X	e.g., regular ‘semi-professional’ participants or

and its people and/or methods					patients in clinical routine
Psychological /physiological state at time of experiment	?	X	X	X	e.g. has the participant recently experienced a traumatic event or consumed psychoactive substances, does s/he have an important appointment later today
Task					
Nature of task	N/A	?	?	?	e.g. speech communication vs. environmental monitoring/detection
Nature of task if speech	X	X	X	X	e.g. repeat, recall, comprehend
Complexity	X	X	?	?	e.g. single vs. multiple tasks
Degree of constraint on route to task fulfilment	?	X	X	X	continuum from e.g. "press the button every time food is mentioned" to e.g. "find out whether you

					have any acquaintances in common".
Exploratory movement	X	X	X	o	degree to which body/head/eye movements by the participant (a) are allowed, and (b) produce realistic changes in the stimuli
Interaction	?	X	X	X	participant as observer/reporter vs. interactor
Predictability	X	X	X	X	e.g. limited response options, pattern of stimuli presentations
Distractors	X	X	?	X	e.g. visuals and audio unrelated to the explicit task
Individual					
Personality	o	X	X	?	e.g. open, agreeable, extroverted, neurotic
Hearing health	X	X	X	X	e.g. type, degree and configuration of hearing loss, tinnitus, hyperacusis

Sensory, cognitive, motor abilities	X	X	X	?	e.g. visual acuity, working memory, balance
Mental health	?	X	X	X	e.g. depressed or anxious.
Competency in task language	X	X	X	o	e.g. native vs. non-native, literacy level
Cultural background	X	X	?	?	e.g. ethnic, socioeconomic or religious factors affecting compliance, social desirability bias
Occupation/ skillsets/ training	?	X	?	X	educational and skill levels and educational attainment
Disease burden	?	X	?	X	e.g. frailty, multimorbidity

Table IV: The independent variables from table IV with examples of design features applicable to each variable that are considered likely to support a high level of ecological validity of a study, and the rating of how well this is currently and generally achieved in clinical and research settings.

Variables (see Table IV for explanatory notes)	**Examples of design features that presumably support a high level of ecological validity**	**Current State of the Art in Typical Studies**		
		Clinic	**Lab**	**Field**
Sources of stimuli				
Characteristic of stimulus sources	The inclusion of varied natural sound sources; non-event speech; different talkers (e.g. male/female, adult/child, native/accent); familiar talkers.	Low	Med	High
Characteristic of stimulus materials	The inclusion of context-dependent cues such as Lombard effects; variation in speed; disfluencies; interjections, and/or emotion.	Low	Med	High

For multimodal stimuli, which modalities carry controlled manipulation	Multiple modalities carry manipulations that are consistent and natural for the intended real-world scenario.	Low	Med	High
How other people are represented	Other people are represented in a manner (e.g. modalities, behavior) that is consistent with the level of realism in other aspects of the scenario's presentation.	Low	High	High
Environment (presentation of stimuli)				
Acoustic field	The presentation of realistic sound levels; spatial relationships; room reverberation.	Med	High	High
Interaction of environment and hearing devices	The acoustic field (including direct and reflected sound) is picked up by the device's microphone/s in a natural manner.	Low	Med	High

Incorporation of dynamic aspects	The presentation of moving sources is realistic for the intended real-world scenario.	Low	Med	High
Modalities included (visual, inertial, etc.)	The presentation includes visual cues (e.g. AV speech cues, non-verbal background cues); tactile cues in interferer stimuli; inertia in the environment.	Low	Med	High
Context of participation				
Participant preparation	Clear instructions and familiarization of study tasks are provided, and the participant understands the purpose of the experiment.	Med	Med	Med
Semantic associations of the situation being simulated for the participant	The situations are familiar and relevant to the participant, and not associated with strong positive or negative affective reactions (unless part of the study design).	Med	Low	High

Motivation to take part	The scenario and task elicit appropriate engagement and motivation, without the need for coercion or undue extrinsic motivation such as disproportionate monetary reward.	Med	Med	High
Familiarity with the lab and its people and/or methods	The participant feels comfortable with physical aspects of the experiment, trusts the personnel, and feels able to withdraw if so desired.	Med	Med	High
Psych/physiological state at time of experiment	The participant is not abnormally stressed or anxious due to factors beyond the study design, and is not inappropriately influenced by drugs, caffeine, etc.	Low	Low	High
Task				
Nature of task	The tasks included are appropriate for the intended real-world scenario.	Med	Med	High

Nature of task if speech (repeat, recall, comprehend, …)	The speech tasks included resemble those that might occur in the intended real-life scenario.	Low	Low	High
Complexity (incl. single vs. multiple tasks)	Any additional tasks included stimulate natural mental processes as they might occur in the intended real-world scenario.	Low	Med	High
Degree of constraint on task fulfilment	The participant is free to perform the task in whatever ways feel natural in the intended real-world scenario.	Low	Low	Med
Exploratory movement	The participant is allowed freedom of gaze, head movement and/or body movement similar to that they would have in the intended real-world scenario, and such movements produce realistic changes in the stimuli.	Low	Med	High
Interaction	Interaction with other persons represented or actually present elicits plausible behaviors from all involved.	Low	Med	High

Predictability	The task possesses predictability similar to what would be present in real life.	Med	Med	High
Distractors	Any distractors are plausible for the intended real-world scenario.	Low	Low	High
Individual				
Variety of personality and demographic factors	Participant recruitment includes stratification or registration of those personal and demographic variables believed to have potential influence.	High	Low	Low