Towards an Optimized Dichotic-Listening Paradigm

Fredrik André Lundgård Samuelsen

MAPSYK360, masterprogram i psykologi,

Studieretning: Atferd og nevrovitenskap

ved

UNIVERSITETET I BERGEN

DET PSYKOLOGISKE FAKULTET

VÅR 2020

Antall ord: 13333

Veileder: Kristiina Kompus, Institutt for biologisk og medisinsk psykologi (IBMP)

Biveileder: René Westerhausen, (1) Psykologisk institutt (PSI), Universitetet i Oslo, (2)
Institutt for biologisk og medisinsk psykologi (IBMP), Universitetet i Bergen

Abstract

Dichotic listening is a well-established non-invasive test, often used to assess hemispheric dominance for speech and language processing. Despite its widespread use in both research and diagnostics, dichotic-listening paradigms appear to have suffered from suboptimal reliability, a problem which undoubtedly threatens the conclusions drawn from obtained measures. In this regard, the present study aimed to design and evaluate a novel dichotic-listening paradigm, optimized for the reliable assessment of hemispheric differences in speech processing. Following an extensive literature review, design features were proposed based on the main experimental factors known to systematically bias task performance or affect random error variance. A central design principle was to reduce task-induced cognitive demand in the attempt to obtain stimulus-driven laterality estimates. The main design features implemented included the utilization of stop-consonant vowel (CV) syllables as stimulus material, a single stimulus-pair in each trial, and a single, free-recall response instruction. Healthy, right-handed young and middle-aged adults (*N* = 50) took part in a test-retest evaluation of a verbal and a manual-response format version of the paradigm. Regardless of response format, intraclass correlation coefficients (ICC) revealed "good" to "excellent" reliability estimates for the full paradigm. The current results indicate that the present paradigm may serve as an effective and efficient alternative to contemporary paradigms both in experimental and clinical settings.

Keywords: dichotic listening, reliability, lateralization, speech perception, neuropsychology

Sammendrag

Dikotisk lytting (DL) er en veletablert non-invasiv test, ofte brukt til å undersøke hemisfærisk dominans for tale og språkprosessering. Til tross for den utbredte bruken både innen forskning og diagnostikk, tydes det på at DL-paradigmer har lidd av middelmådig reliabilitet, et problem som utvilsomt svekker konklusjoner trukket fra testenes målinger. I denne forstand tok den aktuelle studien sikte på å designe og evaluere et nytt DL-paradigme, optimalisert for reliabel vurdering av hemisfæriske forskjeller i taleprosessering. Etter en omfattende litteraturgjennomgang ble designfunksjoner foreslått. Disse var basert på de viktigste eksperimentelle faktorene kjent for å tilføye systematisk bias til oppgaveutførelsen eller for å påvirke tilfeldig feilvarians. Et sentralt designprinsipp var å redusere oppgaveinduserte kognitive krav i forsøket på å undersøke stimulusdrevne lateralitetsestimater. De viktigste designfunksjonene som ble implementert inkluderte bruken av stopp-konsonant vokalstavelser (CV) som stimulansmateriale, et enkelt stimuluspar i hver prøve, og en enkel, fri tilbakekalling som responsinstruksjon. Friske, høyrehendte unge og middelaldrende voksne (*N* = 50) deltok i en test-retestevaluering av en verbal og en manuell responsformatversjon av paradigmet. Uavhengig av responsformat, avslørte intra-klassekorrelasjonskoeffisientene (ICC) «gode» til «utmerkede» reliabilitetsestimater for hele paradigmet. De aktuelle resultatene indikerer at det foreliggende paradigmet kan tilby et effektivt alternativ til eksisterende paradigmer, både i eksperimentelle og kliniske områder.

Nøkkelord: dikotisk lytting, reliabilitet, lateralisering, språkpersepsjon, nevropsykologi

Forord

Denne masteroppgaven ville ikke vært mulig å gjennomføre uten den enorme støtten, entusiasmen og rådgivningen jeg mottok fra Professor René Westerhausen. Allerede fra det øyeblikket jeg meldte ifra om min interesse rundt auditorisk persepsjon, ble jeg tatt godt imot med både åpne hender og ører. Så snart prosjektet ble påbegynt var allerede hypotesen klar, og data fra 17 deltakere var på forhånd blitt samlet inn. Jeg fikk utrolig god opplæring i faget, samtidig som jeg ble tildelt en rekke interessante artikler å lese meg opp på. Alt dette førte til at jeg raskt fikk god innsikt i både den eksperimentelle prosedyren og teorien bak dikotisk lytting. Det tok ikke lange tiden før jeg ble etterlatt på egenhånd som ansvarlig for de resterende eksperimentene i audio-laben på IBMP. Den tilliten og respekten jeg ble møtt med, samt det ansvaret jeg ble tildelt, var deler av et utrolig viktig aspekt i den forstand at prosjektet aldri ble noe jeg kun hadde en assisterende rolle i, men heller noe som sto meg veldig nært. Totalt har prosessen vært en stor fornøyelse, der jeg har fått frie tøyler til å utforske og fundere på diverse forskningsområder innenfor nevropsykologien. Det å kunne ta et dypdykk inn et helt nytt og utrolig spennende felt som dette har vært en av de mest lærerike aktivitetene jeg noensinne har tatt del i. Med dette ønsker jeg å understreke hvor utrolig viktig det er å ha gode veiledere med på laget. Jeg ønsker også å rette en stor takk til Professor Kristiina Kompus som fikk meg i kontakt med René, og som har måttet leve med mine stadig uformelle besøk på kontoret, og ikke minst sine frekvente konsultasjoner med René på vegne av min oppgave. Studien ble finansiert av Psykologisk institutt (PSI) ved Universitetet i Oslo. Takk til Kristin, Nelin og Sarjo for hjelp med datainnsamlingen. Ved hjelp av mine utrolig engasjerte og behjelpelige medstudenter har jeg evnet å opprettholde motet, og jeg er svært takknemlig for deres tilstedeværelse. Ikke minst står jeg i gjeld til alle de tålmodige og forståelsesfulle deltakerne som helt frivillig tok del i mitt eksperiment. Til slutt vil jeg gjerne takke mine nærmeste, spesielt mine foreldre, som har heiet meg frem hele denne veien.

Table of Contents

It has long been acknowledged that the left and the right side (i.e., hemisphere) of the brain differ both in terms of function and structure. With regard to function, each hemisphere appears to be specialized in certain cognitive tasks such as the processing of speech, visuospatial information, and emotion (Hellige, 1993). Such asymmetry is often collectively referred to as “lateralization,” “hemispheric asymmetry,” or “hemispheric specialization.” Furthermore, knowledge regarding hemispheric specialization remains essential in understanding the functional and anatomical properties of the brain. In fact, a variety of neurological conditions has been linked to atypical asymmetry (Rogers, 2014). Thus, research on laterality may provide valuable insight into not only normal but also abnormal brain functioning and anatomy. In order to assess hemispheric specialization, both advanced neuroimaging techniques and behavioral methods can be used. However, behavioral tasks, such as visual half-field experiments and dichotic-listening testing, are less expensive and rather easily administered alternatives as opposed to imaging techniques (Ocklenburg, 2017). Indeed, one of the most frequently used tests for measuring laterality is dichotic listening (Hugdahl, 2016). Although dichotic listening continues to be a mainstay measure of laterality, the question remains whether contemporary testing is being thoroughly planned and executed.

Dichotic listening refers to an experimental technique within the neuropsychological field. The non-invasive methods offered by current dichotic paradigms commonly applies as a way of assessing hemispheric asymmetries in auditory processing. Participants typically equip headphones in which two nearly identical auditory stimuli are presented simultaneously, with one of the stimuli exposed to the left and the other to the right ear (Bryden, 1988). Arguably, this raises a “contest” between the two stimuli, and the “winning” stimulus is of interest to the researcher. Dichotic listening can be experimented with in various ways, with methods ranging from giving the participant special instructions, to manipulating stimuli or context. Inferences drawn from the obtained measures may offer important implications in both research and diagnostics. For instance, clinical evidence has indicated that there is a clear link between the functional properties of callosal fibers and interhemispheric communication during verbal dichotic listening (Westerhausen & Hugdahl, 2008; Musiek & Weihing, 2011). Moreover, dichotic-listening tests have been used excessively as a way of assessing the nature of various disorders and impairments (Wester, Lundervold, Hugdahl, & Taksdal, 1998; Hugdahl, 2000). Nevertheless, the reliability of dichotic paradigms remains suboptimal, potentially threatening conclusions drawn from contemporary findings (Voyer, 1998; Westerhausen, 2019). Knowledge regarding modulating variables of reliability within dichotic-listening paradigms seems to have been, to a certain degree, overlooked by researchers. These variables, however, act as essential utilities in planning, designing, and conducting these experiments, and it is thus crucial to account for their presence. Probably the most widely assessed field of study within dichotic listening as of today regards hemispheric differences within speech and language processing. Hence, the present paper will direct a central focus towards one of the most extensively employed verbal dichotic-listening paradigms and its recognized reliability moderators. This will be done in the ultimate attempt to develop a more reliable dichotic-listening paradigm for the assessment of hemispheric specialization for speech processing.

1.1. Shifting the Focus from Executive Functions to Stimulus-Driven Laterality

The research field surrounding dichotic listening takes hold of theories within a wide range of scientific studies, including neuroscience, psychology, linguistics, and other cognitive sciences. Arguably, the experimental method primarily has its roots within the field of selective attention, a subject of study which originated with the interests of scientists such as Broadbent, Cherry, and Treisman during the renowned cognitive revolution (Hiscock & Kinsbourne, 2011). Briefly, Broadbent argued that in the case of an individual receiving two messages simultaneously, only one of the messages would become processed while the other message would become rejected. That is, he argued that the individual would "filter" out information due to a limited capacity system (Broadbent, 1957). In 1961, however, Doreen Kimura took a step further by introducing these ideas to the neuropsychological field (Kimura, 1961a, 1961b). That is, after Broadbent initiated the first dichotic presentation using digits as stimuli, Kimura adopted the approach and went on to discover profound ear asymmetries in patients suffering from temporal-lobe damage. Around this period, dichotic listening went from originally being considered a pure form of exploring selective attention to becoming one of the leading methods of studying laterality. Although dichotic listening is still being used as a way of assessing selective attention today, these measures presumably reflect a different aspect of lateralization than that of the contemporary verbal dichotic-listening paradigms excluding selective attention. A notable example that might support the notion as to why the study of selective attention within dichotic listening represents other aspects of lateralization lies within the "forced-attention paradigm" (Hugdahl & Andersson, 1986). In such a case, the test subject is instructed to direct attention towards a given stimulus or exclusively to a given ear during the dichotic trial. For instance, the experimenter may request the subject to exclusively repeat stimuli detected on the right ear (i.e., forced-right condition). With the use of such instructions regarding the direction of attention, additional top-down (i.e., controlled) effects are presumed to be supplemented. Furthermore, the ability of an individual to focus attention on task-relevant stimuli and to suppress irrelevant stimuli is assumed to represent top-down modulation (Gazzaley & Nobre, 2012). Notably, Kompus et al. (2012) demonstrated that the use of forced attention makes for a more difficult task for subjects to perform as compared to the application of paradigms allowing for free-selection methods. By the use of a free-recall condition with the dichotic stimuli serving as thoroughly fused percepts, however, it is presumed that top-down effects are minimized and that less cognitively demanding bottom-up (i.e., stimulus-driven) effects are highlighted (e.g., Hiscock & Kinsbourne, 2011). In other words, it is believed that more perceptual aspects of speech processing than that of selective attention may be studied within dichotic listening once the opportunity of the test subject to make preferential choices in detecting stimuli is inhibited – or most favorably, eliminated.

1.2. From Ear to Brain: The Structural Model

Two well-known competing theories are attempting to explain the neural processing of verbal stimuli within dichotic listening as of today: (1) the structural model (Kimura, 1967; see Figure 1) and (2) the attentional model (Kinsbourne, 1970). As the names imply, the structural model sheds light on the interaction between incoming verbal stimuli and neural pathways. In contrast, the attentional model assumes that a pre-stimulus attentional bias occurring unilaterally (i.e., occurring exclusively in one of the two hemispheres) within the auditory cortex (AC) is responsible for eliciting asymmetric processing. With regard to the latter, Kinsbourne's model proposes that the expectancy of incoming verbal stimuli activates the hemisphere of which is dominant for speech processing, which ultimately puts the individual in a preparatory state for the execution of the relevant processing. The structural model, serving as a more anatomically based hypothesis, however, appears to stand out as having received a sufficient amount of empirical support (Westerhausen et al., 2009).

The cerebrum may be divided into two hemispheres—left and right—and these parts interact via the corpus callosum. There is a recurrent finding in dichotic listening that verbal stimuli are within most individuals reported or discovered more rapidly and more precisely in the right ear. This phenomenon is famously known as the right-ear advantage (REA), and it was first discovered by Doreen Kimura (1961a, 1961b). When using consonant-vowel (CV) syllables, REA is typically found in about 85-90% of right-handed individuals and around

65% of left-handed individuals (Hugdahl, 1991). This advantage has been associated with another widely accepted understanding, namely that whereas most non-verbal sounds (e.g., humming and melodies; Kimura, 2011) appear to become processed in the right hemisphere, speech representation is within most individuals found to occur in the left hemisphere (Voyer, 1998; D'Anselmo, Marzoli, & Brancucci. 2016). Kimura already knew from previous research that auditory stimuli tend to be processed contralaterally (i.e., processed through crossed pathways) to the stimulated ear within animals, and thus presumed, following her findings, that this was the same case for humans (Kimura, 2011). More specifically, the theory postulates that stimuli entering the right ear will, in most individuals, be transferred directly to the left hemisphere and that stimuli entering the left ear will be transferred to the right hemisphere, through the corpus callosum and over to the left hemisphere to become processed. In a presumed atypical brain in which left-ear advantage (LEA) is displayed, and speech processing is dominant in the right hemisphere (Hugdahl, 1991), the process is thought to be similar except taking place in the opposite direction. That is, no matter which hemisphere is dominant for speech processing, the contralateral processing of verbal stimuli tends to dominate during the dichotic-listening task (Musiek & Weihing, 2011).

Personally, Kimura believed that the ipsilateral pathways (i.e., pathways at the same side of the hemisphere as in which stimuli enter) occludes during the contralateral processing (Kimura, 2011). Through later dichotic-listening studies on patients with split-brain (i.e., patients who underwent corpus callosotomy or commissurotomy), in which the connection between the two hemispheres has been removed, findings have been in favor of this occlusion theory. For instance, it has been found that these patients are able to easily perceive monaural stimuli (i.e., exposure to stimulus on an exclusive ear) from both ears. However, during dichotic listening, they struggle to detect stimulus on the left ear as opposed to the right ear, which displays a REA (Westerhausen & Hugdahl, 2008; Hiscock & Kinsbourne, 2011).

The argument as to why these patients are able to perceive stimuli from the right ear—but not from the left ear—has been attributed to the basic mechanisms of auditory perception. The auditory nerve goes from the cochlea down to the brain stem. The incoming auditory stimulus is then transferred and analyzed through the cochlear nucleus, superior olivary complex, lateral lemniscus, and inferior colliculus before it both contralaterally and ipsilaterally reaches the primary auditory cortex via the medial geniculate of the thalamus (Plack, 2004). The contralateral projections, however, are presumed to be dominating, and that is why it may be the case that stimuli entering the right ear reaches the left hemisphere—which in most individuals appears to be dominant for speech processing—more rapidly than stimuli entering the left ear (Hugdahl, 2003; Westerhausen & Hugdahl, 2008).

As seen earlier, verbal stimuli entering the left ear is within most presumably required to be transferred via corpus callosum from the right to the left temporal lobe before it may undergo the relevant speech processing. This additional detour, which supposedly causes a considerable delay and presumably plays a significant role in the origins of the REA has, by Zaidel (1983), been coined as the "callosal relay" model. This model, stressing the importance of considering callosal properties, can be interpreted as an extension of the structural model. Namely, the callosal relay model makes further assumptions by hailing to the idea that the magnitude of the ear advantage is dependent on the functional characteristics of the corpus callosum. Indeed, it has been demonstrated that the interhemispheric transmission of auditory information appears to be susceptible to callosal lesions and abnormalities (Westerhausen & Hugdahl, 2008), and even to the developmental course of the corpus callosum (Musiek & Weihing, 2011). The latter in line with the observation of a REA increase with age (Strouse Carter & Wilson, 2001; Westerhausen, Bless, & Kompus, 2015). Thus, obtained measures from the verbal dichotic-listening task not only reveal hemispheric asymmetries but may also serve as an indicator of interhemispheric connectivity and, ultimately, callosal functionality.

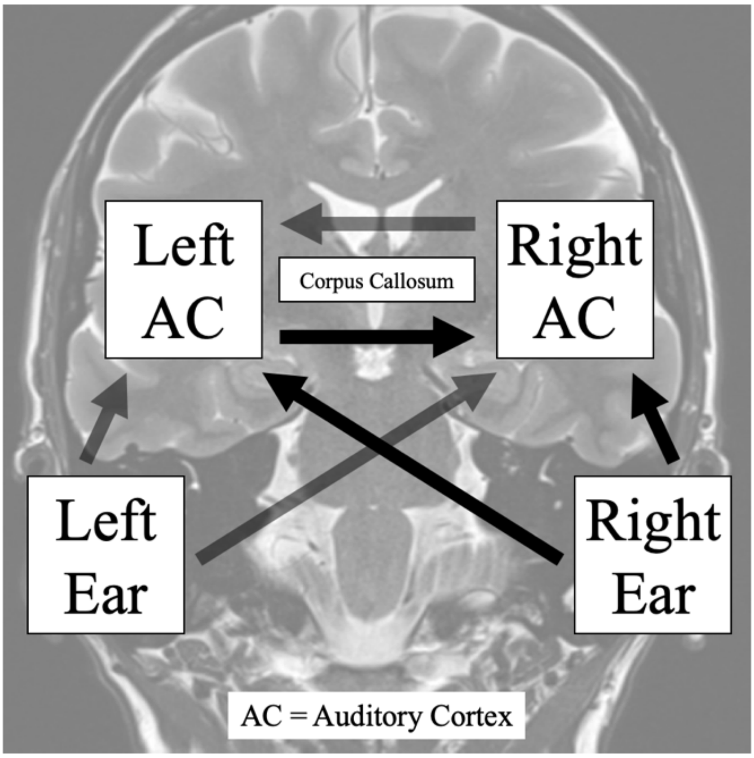

Figure 1. Schematic representation of Doreen Kimura's structural model of dichotic listening.

1.2.1. Expanding the dichotic-listening theory. Although Kimura's structural model may have received the most support, alternative models have later been proposed with both alterations to the initial model and different perspectives on the source of the ear advantage. Although many of these theories are mostly in disagreement, it is universally agreed upon that, in most, the right ear has a preference for verbal material, and that this reflects a left-hemispheric dominance for speech processing (Westerhausen & Kompus, 2018). Probably the most contrasting theory to the structural model is Kinsbourne's attentional model (Kinsbourne, 1970), a view which only received a limited amount of support in subsequent studies (Pollmann, 2010). However, more recently, so-called "two-stage models" have also been suggested (Hiscock & Kinsbourne, 2011), including "signal detection versus localization" (e.g., Hiscock et al. 1985; Hiscock, Inch, & Ewing, 2005) and "bottom-up versus top-down processing" (e.g., Westerhausen & Hugdahl, 2008; Hugdahl et al., 2009; Westerhausen et al., 2009; Westerhausen et al., 2015; Hugdahl, 2016; Westerhausen &

Kompus, 2018). Both of these models assume that bottom-up and top-down processes interact during dichotic listening. However, the models disagree concerning top-down processing, in which the “Hiscock and Kinsbourne” model makes assumptions based on the less-supported idea of a left-hemispheric attentional bias as initially proposed by Kinsbourne (1970). In contrast, the “Hugdahl et al.” model follows the basic principles of the structural model.

As principally inspired by the ideas and contributions of Phil Bryden (Bryden, Munhall, & Allard, 1983; see Hugdahl, 2016 for review), the fundamental principle of the “Hugdahl et al.” model is that the observed ear advantage in a dichotic trial—of which stimulus presentation has been optimized for the studied subject—is thought to be dependent on two information-processing stages. Moreover, dichotic listening can be seen as including (1) an initial bottom-up (stimulus-driven) stage and (2) an additional top-down (instruction-driven) stage (Westerhausen & Kompus, 2018). Furthermore, a perceptual representation of the dichotic stimuli is at first assumed to be established in the auditory short-term or working memory. Here, they remain prone to a second-stage modulation whereby cognitive-control processes seek to select a response in line with the given task instructions. In the first, bottom-up stage, it is believed that the more salient stimulus will “win.” That is, if the stimulus features and presentation allow for an uneven perception of the two stimuli, the stimulus of which is perceived as dominating will become processed. Importantly, the deciding factor as to what makes one of the two stimuli favorable in terms of perception can be influenced by several variables. During the second, top-down stage, however, task instructions may engage executive functions in the attempt to modulate the initial processing. In other words, the stimulus that initially “won” can potentially become replaced by the opposite-ear stimulus, constituting a competition. Thus, task instructions (e.g., left-directed attention) discordant with the initially favored stimulus (e.g., right-ear stimulus) involves a processing conflict that needs to be overcome, requiring an executive cognitive control process (Hugdahl et al., 2009).

The current "Hugdahl et al." model is grounded in decades of research. Perhaps the most direct evidence for this theory can be found within the forced-attention paradigm, which typically reveals a clear REA in the non-forced condition (i.e., free-recall instruction), with a stronger REA in the forced-right condition, and an obvious shift towards a LEA in the forced-left condition (Hugdahl et al., 2009; Hugdahl, 2011; Hugdahl, 2016). Furthermore, a globally undeniable observation during dichotic experiments is that even individuals who reveal an overall REA report substantial amounts of the stimuli presented to the left ear (Westerhausen & Kompus, 2018). Findings such as these seem to suggest that the ear advantage observed in the dichotic-listening task must have been contributed to by external factors, rather than having been solely determined by a presumed, innate hemispheric specialization. Additionally, it has been demonstrated with dichotic listening that about 20% of right-handed individuals display a presumed atypical right-hemispheric specialization by revealing a LEA (Bryden, 1988). However, more direct measures such as fMRI and the Wada-test have yielded less prevalent estimates of atypical laterality of 5% (Badzakova-Trajkov, Häberling, Roberts & Corballis, 2010) and 13% (Carey & Johnstone, 2014) within the same population. Thus, seeing that the measures obtained from the dichotic experiment are inconsistent, the ear advantage must be susceptible to non-hemispheric sources of error.

The degree to which measurements can be replicated is referred to as reliability (Koo & Li, 2016). Given that the ear advantage appears to be sensitive to factors other than the "inherent" laterality, the reproducibility of the measures may be threatened by unseen sources of modulation. Along with stimulus presentation features, top-down modulatory factors such as attention and cognitive control are thus presumed to systematically affect the reliability of verbal dichotic-listening paradigms. Hence, the present paper will aim to exclusively pursue "pure" stimulus-driven laterality by attempting to limit the sources that allow for top-down modulation, and to control for design features that may bias the initial stimulus selection.

1.3. Design Features and the Potential for Increasing Reliability

1.3.1. Stimulus material and the number of trials. The stimulus material used within this field of study can be divided into three main categories: numerical words, non-numerical words, and non-word syllables. Digit pairs, mono- or bisyllabic substantives, fused dichotic rhyming words (FDRW), consonant-vowel-consonant (CVC), stop consonant-vowel (CV), and vowel-consonant-vowel (VCV) have so far reigned as stimulus material (Westerhausen, 2019). All variations of stimulus material, as mentioned above, have been directly validated towards indications of hemispheric asymmetry. This has been done through, for instance, the Wada-test (i.e., a "selective injection of a barbiturate into the right or left hemisphere, "silencing" the injected hemisphere for about 10 min"; Hugdahl, 1991, p. 24), performance deficits due to lesions, or hemispherectomy (i.e., a surgical intervention in which one of the two hemispheres are removed) (Westerhausen, 2019). Put another way, each of these stimulus materials indicates "construct validity." That is, they measure what they are intended to measure (Pedhazur & Schmelkin, 2013).

A major difference between these categories of stimulus material, however, is that they are presumably measuring different aspects of speech processing. Verbal stimuli introduce a wide variety of factors ensuring that the acoustic speech input that is perceived by the ear may elicit varying conceptual and semantic representations in the brain. Factors such as distinctive features, phonemes, syllable structure, phonological word forms, grammatical features, and semantic meaning can seemingly be divided into different levels of analysis (Hickok & Poeppel, 2007). Hence, it does not seem implausible to speculate on whether non-word syllables—serving as "building blocks" for meaningful words—may take part in an earlier processing stage than numeric words and non-numeric words (Westerhausen, 2019).

The most commonly used non-word syllable version as of today, CV, bases syllables on six plosive consonants (/b/, /d/, /g/, /p/, /t/, /k/) combined with a vowel (usually /a/). With each of the six syllables combined, a total of 36 pairings become available. Both the forced-attention and free-recall instructional conditions have been utilized interchangeably in conjunction with the CV-version (Hugdahl, 2003). With regard to reliability, the CV has proven to display higher values than that of numeric and non-numeric stimulus material (Voyer, 1998). However, the typical CV-paradigm practiced today usually consists of about 30 trials and tends to reveal “moderate” reliability surrounding $r = .61$ (Westerhausen, 2019). A study that is believed to display “good” reliability, on the other hand, should preferably exceed $r = .75$ (Koo & Li, 2016). By comparing published reliability data from different dichotic-listening studies, Westerhausen (2019) found that satisfactory reliability values of $r > .80$ appear to be found when around 120 trials or more are being used. It is worthy of note, however, that the use of more than 120 trials may eventually pose a threat to the feasibility of the experiment. That is, the length of the experiment would not seem likely to last any longer than 10 minutes with the employment of 120 trials. Nevertheless, while the use of more than 120 trials might increase the overall reliability, it carries the risk that the experiment will simply become lengthy for potential participants. Thus, 120 trials appear to be ideal.

1.3.2. Other modulating variables of reliability. Notably, reliability is not a measure that is solely influenced by the stimulus material nor by the number of trials. In his comprehensive review of dichotic-listening experiments with a focus on lateralization within speech processing, Westerhausen (2019) listed five design principles for the planning and execution of a reliable dichotic-listening paradigm (see Table 1; Westerhausen, 2019, p. 763).

Table 1

Five design principles for a reliable dichotic-listening paradigm, measures how to adhere to these principles, and benefits achieved

Principle	Measures	Benefit
1 **Keep it simple**	Single stimulus pair per trial, single response, immediate response collection, free-recall instruction	Minimizes working-memory load and the demand for cognitive-control process
2 **Maximize the spectral and temporal overlap of the paired stimuli**	Use rhyming stimuli, assure inter-channel onset synchronization	Increases the likelihood of perceptual fusion, reducing the cognitive demands of the paradigm
3 **Equalize perceptual difficulty of the paired stimulus**	Comparable stimulus difficulty and perceptual saliency, control for negative priming between trials, present all stimuli equally often to the left and the right ear	Prevents or controls for processing biases favoring one ear or the other stimulus
4 **Ensure fair testing conditions**	Hearing acuity testing, good testing environment, language appropriateness	Minimizes the likelihood that factors other than perceptual laterality affect the measures
5 **Collect a sufficient amount of responses**	Use a minimum number of 90 (dichotic) trials, preferably 120	Improves ratio of effect variance to random error variance

Note. Adapted from «A primer on dichotic listening as a paradigm for the assessment of hemispheric asymmetry,» by Westerhausen, R., 2019, *Laterality: Asymmetries of Body, Brain and Cognition, 24*(6), p. 763.

Each of the individual design principles and their related measures from Table 1 (Westerhausen, 2019, p. 763) are based on modulating variables for the reliability of perceptual laterality measures in dichotic-listening paradigms (see Westerhausen, 2019 for review). Firstly, the remainder of the identified reliability moderators and suggestions on how to control them will be addressed. Secondly, the five principles taken from this table will be adhered to in order to evaluate their functional outcome with regard to reliability. This evaluation will be executed by testing proposed design features that implement the suggested design principles with the currently applied methods of the standard CV-paradigm.

The number of stimulus pairs and responses per trial. Throughout the years, both multiple (Kimura, 1961a, 1961b) and single (Wexler & Halwes, 1983) stimulus-pair trials have been employed in dichotic-listening experiments. In the first dichotic study conducted by Kimura, for instance, three dichotic pairs of digits were presented in close succession before the participant was instructed to report back. However, presenting multiple stimulus pairs as compared to single stimulus pairs per trial appears to increase working-memory load, resulting in an amplified REA (Penner, Schläfli, Opwis, & Hugdahl, 2009). Namely, with an increasing demand on cognitive resources as brought about by successive pairs, modulatory effects on the REA seems to occur. Thus, by accounting for the possibility that the lack of correct left ear recalls may be owed to the administration of multiple-stimulus pairs, single-stimulus pairs appear to be the optimal alternative for "pure," bottom-up laterality testing.

Whereas multiple-stimulus pair trials require multiple responses, single-pair trials do not usually require more than a single response. However, subjects have also been instructed to report back both stimuli in single-pair trials as well (Studdert-Kennedy, Shankweiler, & Schulman, 1970). Multiple-pairs introduce a longer stimulus-to-response delay so as to allow for multiple answers. When non-interfered—such as in the case of a single-pair trial—response delay might seem beneficial as the REA has been found to increase in magnitude with short—one- or three-second-long—extensions of an immediate retention interval (D'Anselmo et al., 2016). However, interpreting stimulus-to-response delay in the context of rehearsal effects and working-memory involvement (Baddeley, 2012), increased response delay arguably plays a major role in magnifying the ear advantage (Voyer, Dempsey & Harding, 2014). That is, the selected stimulus might be stored in the phonological working memory, allowing for a stronger representation, which subsequently exaggerates the resulting response. For that reason, it appears that immediate response would be the most qualified procedure to administer in single-pair paradigms aiming to investigate structural laterality.

Stimuli alignment. Along with longer stimulus-to-response delays, interaural asynchrony in stimuli-onset may also lead to the engagement of higher cognitive functions. Even onset delays of the left-ear stimulus as short as 20-30 ms behind the right-ear stimulus have been found to generate a LEA (Studdert-Kennedy et al., 1970). Onset asynchrony, however, is not the sole stimuli-alignment factor affecting ear preference. Even when temporally aligned in terms of both onset and offset, the unvoiced syllable /pa/ and the voiced syllable /ba/ would, for instance, differ in their spectro-temporal properties. This spectro-temporal asynchrony can be attributed to dissimilarities in determinants such as the first formant-onset time, voice-onset time (VOT), and voice length of the syllables. Although lacking a perfect overlap—which would, in any event, be an unnatural feature—CV-syllables has the advantage of differing in nothing but a single initial phoneme and thus qualifies for maximizing the spectro-temporal overlap (Repp, 1977). In other words, perceptual fusion is likely to be accomplished within the CV-paradigm. With the use of fused percepts, the individual may experience the advantage of exclusively perceiving one of the simultaneous stimuli during a trial in the dichotic-listening task (Techentin & Voyer, 2011; Westerhausen, Passow, & Kompus, 2013). When the paired stimuli overlap both spectrally and temporally as such, the dichotic task might be regarded as a less demanding situation. Likewise, by comparing the effects of fusing and non-fusing stimulus pairs using fMRI, Westerhausen et al. (2013) demonstrated that fusing stimuli benefit from shorter response times and reduced activity in inferior-frontal regions. Furthermore, the authors additionally put forth that different cognitive processes contribute to the ear advantage depending on the degree to which the stimuli fuses. More specifically, an increase in reactive cognitive control (i.e., the involvement of modulatory attentional control after cognitively demanding events; Braver, Paxton, Locke, & Barch, 2009) was observed with pairs of lower overlap, as opposed to stimulus pairs with high spectro-temporal overlap (Westerhausen et al., 2013).

An important and often overlooked determinant for the likeliness of accomplishing fusion even in paradigms such as the CV, however, is the voicing category of the paired stimuli (e.g., Gadea, Gomez, & Espert, 2000; Dos Santos Sequeira et al., 2010; D'Anselmo et al., 2016). The voicing category of paired CV-syllables can be divided into voiced (e.g., /ba/ or /da/), unvoiced (e.g., /pa/ or /ka/) and mixed syllables (e.g., /ba/ presented together with /pa/). Voiced syllables are characterized by their short VOTs (in Norwegian 20-30 ms; see Rimol, Eichele, & Hugdahl, 2006), whereas unvoiced syllables contain longer VOTs (in general 50-83 ms; see Westerhausen, 2019). Moreover, VOT refers to the time point in which the vocal cords start to vibrate following the initial release of the consonant burst (Voyer & Techentin, 2009; Westerhausen & Kompus, 2018). It has been demonstrated that, with the use of a mixed-voicing category within a trial (e.g., a combination of /pa/ and /ba/), perceptual fusion is less likely to occur, and laterality indices are subjected to modulation. Precisely, when paired with a voiced stimulus, the unvoiced stimulus almost exclusively determines the ear advantage, irrespective of which ear it is presented to (Speaks, Niccum, Carney, & Johnson, 1981; Rimol et al., 2006; Voyer & Techentin, 2009). That is, having an unvoiced syllable (e.g., /ka/) presented to the left ear and a voiced (e.g., /da/) to the right would result in a LEA. Presenting the unvoiced syllable to the right ear and the voiced to the left, on the contrary, would reveal a REA. To control for such dominance effects–which appear during the pairing of syllables with different VOTs–it is thus believed that it would be beneficial to avoid the application of mixed-voicing pair trials. Rather, pairing syllables of the same-voicing category only (i.e., (1) short-short and (2) long-long) appears to be the optimal solution. An example of the temporal difference in the VOTs of a voiced and an unvoiced syllable is illustrated in Figure 2.

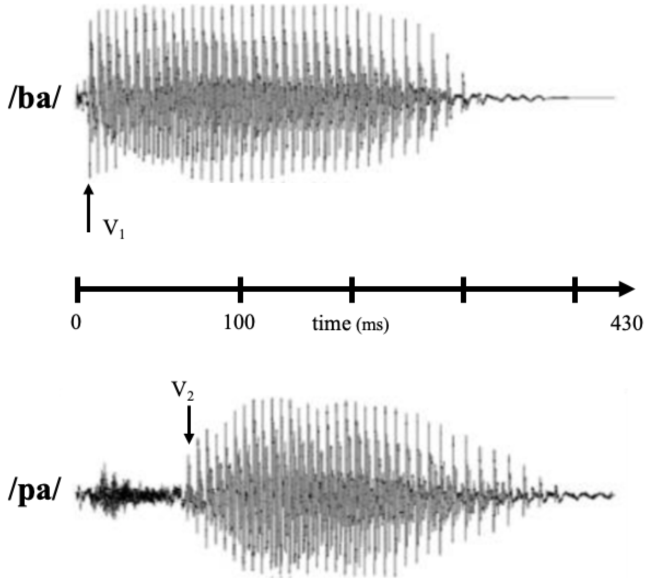

Figure 2. Waveform examples of the voiced syllable /ba/ and the unvoiced syllable /pa/. Timepoint "0" refers to the stimulus onset. V_1 and V_2 refer to the approximate voice onsets within the two syllables.

Stimulus difficulty and trial order. Ideally, all forms of biases favoring a given stimulus or one ear should be avoided. During competitive trials, variations in perceived item difficulty may determine the exclusive attentiveness towards specific speech stimuli. Within languages, for instance, particular words occur more often than others, and these words occurring in a higher frequency are known to be processed more rapidly and to be recollected better than that of infrequent words (Brysbaert, Mandera & Keuleers, 2018). It appears to be the same case for non-word syllables. In Italian subjects, for example, a stronger tendency to report back non-words in which the initial syllable is of high frequency has been reported (Tremblay, Deschamps, Baroni, & Hasson, 2016). Notably, the term "frequency" as used here refers to word occurrence rates and is not to be mixed with the term of musical frequency.

The frequency of words and syllables, however, does not determine the lexical difficulty of speech recognition alone. When presented with speech stimuli, the phonological input of a given stimulus activates memory representations that resemble that of a range of other inputs. More specifically, the Neighborhood Activation Model postulates that a word is harder to identify when the number of phonologically similar words stored in the mental lexicon is high (i.e., when the word is derived from a high-density neighborhood) (Luce & Pisoni, 1998). By utilizing dichotic stimuli based on rankings of word frequency and neighborhood density, Strouse Carter & Wilson (2001) presented “easy” words (i.e., high frequency, low density; e.g., lamp) and “hard” words (i.e., low frequency, high density; e.g., clam) to listeners. They found that the most accurate reports consisted of “easy” words, more precisely when they were paired with “hard” words. That is, presenting a “hard” word to the right ear and an “easy” word to the left revealed a small, yet non-significant, LEA. A REA was revealed with the stimuli presented in the opposite direction, emerging as significantly increased in contrast to trials that included words of equal difficulty (e.g., “hard”-“hard”).

Thus, it appears that the lexical difficulty of the stimuli, both in terms of word frequency and neighborhood density, needs to be accounted for when planning the dichotic experiment. The most favorable alternative occurs to encourage a matching of the two stimuli in each trial for lexical difficulty. Considering that the CV-paradigm has a strictly limited selection of stimulus material, however, this type of matching appears to be a difficult task to perform. Instead, averaging the imbalanced item difficulties across all trials appears to be a fruitful approach. That is, to present all of the different stimulus pairs in a perfectly balanced frequency to both orientations - i.e., left-right ear and right-left ear orientation.

The presentation order of the different stimulus pairs is not to be considered straightforward, however. It has been revealed that the use of trial orders in which stimuli are repeated from a given trial to the next, there is a reduced chance for the repeated stimuli to be reported again (Sætrevik & Hugdahl, 2007a, 2007b). To avoid this effect of negative priming, the order of the trials should be thoroughly considered. While holding on to the idea of utilizing same-voicing pairs only, it only appears appropriate to adopt a pseudorandomized order presenting alternate voiced (“short-short”; VOT) and unvoiced (“long-long”) pair trials.

Fair testing conditions. Finally, experimenters should account for the fact that languages consist of their own characteristic set of phonemes, with syllables having a language-specific structure. In the process of incorporating sounds with meaning, these two factors, including the language-specific distinctive features of speech—such as the difference in how a Norwegian perceives and pronounces /r/ as compared to an American—are essential for the execution of phonological analysis (Hickock and Poeppel, 2007). In fact, linguistic dissimilarities of test subjects may introduce perceptual biases. For instance, Gathercole (1995) demonstrated that the recall accuracy of non-words is greater when non-words resemble familiar words. By employing non-word syllables that are recorded in an unfamiliar language to the test subject, other factors than perceptual laterality may thus be speculated to affect the measures. Hence, stimuli should be language-relevant to all subjects, and the inclusion criteria should ensure that the test subjects preferably have the same or at least a similar linguistic basis. In controlling for the appropriateness of the employed stimuli, the experiment may include homonymic trials. That is, presenting identical stimuli to both ears simultaneously for control.

1.4. A New Paradigm

In summary, the presently reviewed material points to the potential for developing a more reliable dichotic-listening paradigm. The typical CV-paradigm tends to generate indications of mediocre reliability, which ultimately affects the conclusions drawn. Estimates of reliability may be influenced by several factors, including stimulus characteristics, the setup of trials, and instructional conditions (Westerhausen, 2019). By implementing modifications derived from the here reviewed literature to the common design of the CV-paradigm, a new and potentially more reliable paradigm emerges. The main design features of the new paradigm are summarized in Table 2. A detailed description of the experimental procedure, hearing acuity testing, and testing environment is available in the Method section.

Table 2

Design features of the here suggested, new dichotic-listening paradigm and arguments for their implementation

Design feature	Argumentation
1 Stop consonant-vowel (CV) syllables as stimulus material	Proven to be valid test material; CV reveals higher reliability than numeric and non-numeric words as stimulus material (Voyer, 1998)
2 Pairing CV stimuli from the same-voicing category only	Same-voicing category pairs are more likely to fuse into one percept than mixed pairs; reduces the relevance of attention and cognitive-control processes (Westerhausen et al., 2013)
3 Alternating trials of voiced and unvoiced stimulus pairs	Limits negative-priming effects (Sætrevik & Hugdahl, 2007a, 2007b) by preventing stimulus repetition between consecutive trials
4 Free-recall instruction	Reduces task difficulty and relevance of cognitive-control processes compared to selective attention instructions (Kompus et al., 2012)
5 Single-stimulus pair per trial; single, immediate response	Minimizes working-memory load compared to multi-stimulus trials and delayed response paradigms (Penner et al., 2009)
6 Paradigm length of 120 dichotic trials	Previous studies indicate reliability estimates > .80 mostly for paradigms using 120 trials (Westerhausen, 2019)
7 All stimulus pairs are presented in both orientations (i.e., left-right ear and right-left orientation) in an identical frequency	Averages otherwise uncontrolled biases across these trials (e.g., item difficulty effects; see Strouse Carter & Wilson, 2001)
8 Includes binaural (homonymic) trials	Allows to demonstrate stimulus appropriateness, i.e., whether the subject can identify the used stimulus material (e.g., see Gathercole, 1995)

It is here suggested that the design of a verbal dichotic-listening paradigm accounting for each of the recognized modulating variables for reliability, will ultimately make for effective and efficient testing of hemispheric dominance for speech processing. The present paper aims to answer whether the here-suggested, new, and potentially optimized paradigm design will provide higher reliability estimates than that of the presumed “good”-reliability limit, $r = .75$. By attempting to minimize the relevance of higher cognitive functions and its associated error variance, it is predicted that the estimates will exceed $r = .75$. The paper also introduces an additional secondary hypothesis in which it is predicted that the reliability will exceed $r = .75$ for two different response modes; manual and verbal. Namely, the type of response mode applied in a dichotic-listening experiment may influence the outcome in terms of the varied cognitive demands introduced by the manual- and the verbal-response modes. If the primary measured outcome is accuracy measures, however, it is believed that both manual and verbal responses should equally be a prime fit for the study (Westerhausen, 2019).

2. Method

2.1. Participants

The final sample consisted of $N = 50$ right-handed adult university students and employees ($n = 30$ female, 60%, $M_{\text{age}} = 24.98$, $SD_{\text{age}} = 4.16$, age range: 19-39 years). Data from a group of participants ($n = 17$) was collected by René Westerhausen at the University of Oslo. An additional group of participants ($n = 33$) was subsequently recruited at the University of Bergen. Participants were fluent speakers of the Norwegian language (42 native speakers, eight non-native speakers). A high identification rate ($M = 98.3\%$, $SD = 2.2\%$) of the used stimuli when presented in homonymic trials guaranteed that the stimulus material was fitting for all participants. Participants were required to be right-handed (as assessed with Edinburgh Handedness Inventory; Oldfield, 1971; see Appendix C), healthy, without a history of psychiatric or neurological conditions (as assessed by oral self-report before the experiment began) and had to fulfill requirements for normal hearing. The requirement of being right-handed was based on previous research suggesting that left-handed individuals tend to display more atypical patterns in lateralization than that of right-handed individuals (e.g., Westerhausen et al., 2015). The final sample excluded two participants in which one participant only identified 59 of the maximum 72 homonyms correctly (81.9%), and one did not meet the hearing acuity requirements (see Measures and Procedure for details). The recruitment process included the use of posters (see Appendix A) and the researcher's network. The study was announced as a completely voluntary experiment on cognition with information regarding the right-handedness requirement, duration, location, compensation, and anonymization of data.

2.2. Ethical Concerns

The present study has been reviewed by and received ethical clearance through the Department of Psychology's Research Ethics Committee, University of Oslo (Ref.: 3838804). Before each experiment, all participants were handed informed consent (see Appendix B) that had to be read, accepted, and signed. All study data has been saved, analyzed, and published with regard to anonymity. Participants who did not fulfill the inclusion criteria for this study were excluded from the analysis. The anonymous data files from E-Prime (Psychology Software Tools, Pittsburgh, PA) were saved to the personal computer used for the experiment in the audio lab at the Institute for Biological and Medical Psychology (IBMP). After each participation, all data files were backed up on a private and safely kept USB-stick. Participants were anonymized as soon as each experiment was completed. Data files were labeled by an anonymous ID-number representing the participants, and each participant generated two data files differentiating between the manual and verbal response mode sessions. The two data files representing each participant were thus organized by the given ID-number and the session number (e.g., 350-1 and 350-2). Contact details were not saved. That is, it would not be possible to link the collected data to the participant. Participation was completely voluntary, and the participant had the right to withdraw from the study at any time without the need for reasoning nor any form of punishment. Participants were able to request the removal of the data at any time before anonymization. Given that dichotic listening is a non-invasive method that should not require more than the cognitive capacity to perceive, sustain attention, execute cognitive control, to express oneself verbally and to coordinate finger movements, it has been considered to be of low risk for the participant to experience any particular form of stress. In the worst-case scenario, the participant may have experienced moments of drowsiness during the experiment.

Following each session, the participant received a 150,- Norwegian kroner universal gift card (http://www.presentkort.no) as compensation for the study participation. Ahead of departure, the participant was asked to sign a confirmation on receiving the gift card. This confirmation, as well as the informed consent, was stored in a safe in the supervisor's office.

2.3. Measures and Procedure

Hearing acuity testing using an audiometric screening program—Oscilla AudioConsole (Natus Medical Incorporated, Pleasanton, California)—headphones (Sennheiser HD280) and a handheld response button was completed before the dichotic study. Substantial differences in interaural acuity systematically affect the direction of the ear advantage in that the more "well-equipped" ear persistently dominates perception (Speaks, Blecha, & Schilling, 1980). Thus, to be included in the data analysis, the participant had to be able to refer to an average interaural (i.e., left-to-right ear) hearing difference within the range of ±10 dB, and hearing impairments of more than 20 dB SPL for any frequency was unacceptable (Hugdahl et al., 2009). This was tested through a presentation of sine wave tones with the following frequencies: 250, 500, 1000, 2000, and 3000 Hz. The participants were informed that the audio signals would appear at various decibel levels and that the tones would appear firstly at the right ear and then at the left ear midway through the test. Participants were also instructed to press the response button as soon as a sound was detected. The screening took about four minutes in total, and the results of the hearing measurement were recorded on paper. All participants revealed an average interaural threshold difference within the range of ±10 dB (ranging from 8 to -8 dB in favor of the left and right ear, respectively; *M* = -0.64, *SD* = 4.07) across the tested frequencies.

The DL-experiment was conducted in a quiet test room, using E-Prime (Version 3.0 at the University of Oslo; Version 2.0 at the University of Bergen; Psychology Software Tools, Pittsburgh, PA). The use of a quiet test room was an important methodological consideration, considering that the REA is prone to a reduction in the face of background noise (Dos Santos Sequeira et al., 2010). Ahead of the experiment, participants were briefly introduced to the stimulus material. Each participant was instructed to immediately report back what was "heard best" (Bryden, 1988). The participants were at first exposed to a short test block—including ten trials—for familiarizing with the setup and procedure, and then six subsequent blocks with "real" trials. The syllables were presented in the form of recordings of an adult Norwegian male voice actor, keeping intensity and intonation constant. This was done in an attempt to accomplish perceptual fusion (see Westerhausen, 2019). Sound pressure levels of stimuli were confirmed to be sufficient by each participant (as assessed by self-report), indicating a satisfactory sensation level (dB SL). That is, a set decibel level was administered in which each stimulus proved intelligible to each participant (Musiek & Weihing, 2011). Each trial was 4000 ms long. Included in each trial were a preparation interval (1000 ms), stimulus presentation (500 ms), and response-collection interval (2500 ms). The asynchrony of stimulus-onset was set to 4000 ms (Westerhausen et al., 2013). There was room for short breaks between each block (self-administered), and the entire study took approximately 45 minutes to complete. In each block, there were 46 pseudorandomized trials (i.e., combinations of CV stimuli). These 46 trials consisted of a total of 40 dichotic pairs of stimuli and six homonyms for control. Thus, the maximum correct score available for each block was 40 (excluding the six homonyms). Each trial included a set of two CV-stimuli—one on each ear—in which the participant was exposed to a combination of either /ba/, /da/, /ga/, /pa/, /ta/, or /ka/. Every other trial consisted of (1) a combination of voiced (e.g., /ba/-/da/) and (2) a mix of unvoiced syllables (e.g., /pa/-/ka/) (Sætrevik & Hugdahl, 2007a, 2007b). All syllables were presented 46 times each. In most trials, stimuli on each ear were different (i.e., dichotic) except random cases in which stimuli were homonyms for control. To assess test-retest reliability, the study—with exception from the test blocks—consisted of 3x46 trials completed twice for both manual and verbal response mode. In total—accounting for both response modes—there were two familiarization blocks and 12 "real" blocks, altogether consisting of 572 trials.

Responses were collected in an order following a balanced AABB design. That is, every other participant was instructed to begin with responding either manually (*n* = 24) or verbally (*n* = 26). Specifically, if the first participant were to start the experiment by responding manually, the second participant would begin with responding verbally. If the participant was instructed to respond manually, the participant had to press the keyboard buttons on the number pad (numbers 1 to 6) marked with the different stimulus options, based on the syllable perceived during the trial. With the verbal response, however, the participant was instructed to repeat the perceived syllables to the experimenter. The experimenter, who was consistently blind to the syllables presented, logged the verbal response in the same way as the participant did during the manual response. During stimulus presentation, a fixation cross (+) was displayed in the center of the PC-monitor. As soon as a response was logged, the cross got replaced by a circle (o) confirming registration of the response. Correcting an already registered response was not made possible.

Both sites (Oslo and Bergen) utilized the same setup and procedure. Identical audiometer setups, the same headphones models, and quiet test rooms were used. However, there were unavoidable differences in the exact equipment (i.e., computer build and keyboard). Although E-Prime (Psychology Software Tools, Pittsburgh, PA) was used in both labs, version 3.0 was available in Oslo, whereas Bergen was equipped with version 2.0. Nonetheless, considering that the effects of interest in the present study are based on comparisons within participants and accuracy measures rather than the exact reaction times, it is not believed that these differences had a relevant effect on the present results.

2.4. Data Preparation

In preparation for data analysis, the raw data from E-Prime (Version 3.0 at the University of Oslo; Version 2.0 at the University of Bergen; Psychology Software Tools, Pittsburgh, PA) was re-coded in MATLAB (Version R2019a; The MathWorks Inc., Natick, Massachusetts). The two data files generated from each participant were merged into a single MS Excel (Version 16.32; Microsoft Corporation, Redmond, WA) spreadsheet containing individual representative rows for each participant. The test blocks for familiarizing and other unnecessary data (e.g., date and time) were removed from the data during this process. The final re-coded participant data was ranked by ID, presenting information regarding the first response mode given (either manual or verbal), the amount of correctly reported homonyms for both response modes, and the amount of correctly reported dichotic stimuli from each ear (left and right) within both response modes. The latter information was presented as the correctly reported amount in each of the six blocks within both response modes. The spreadsheet was imported into SPSS (Version 25; IBM Corp., Armonk, NY) for analysis.

2.5. Statistical Analysis

The relative difference between the amount of correctly recalled left- (L_c) and right-ear (R_c) stimuli was determined as laterality index (i.e., LI; $LI = (R_c - L_c) / (R_c + L_c) \times 100$) per person and test run (Westerhausen, 2019). The LI, yielding values between +100 and -100, was calculated so as to quantify the magnitude of the ear advantage (Hugdahl & Hammar, 1997). Moreover, positive values (1 to 100) refers to a REA, negative values (-1 to -100) resemble a LEA, and values of zero indicate a no-ear advantage (NEA). Test-retest reliability was estimated separately for the LI, L_c, and R_c, as obtained from the manual- and the verbal-response format paradigms by using intraclass correlations (ICC; correlation coefficients reported as r_{ICC}). A warning has been issued regarding the use of test-retest reliability in that low estimates may be representative of low reliability, changes in the test subject, or both (Pedhazur & Schmelkin, 2013). The ICC, however, serves as a reliability index reflecting the extent to which measurements both correlate and agree (Koo & Li, 2016). A two-way mixed-effects model aiming for absolute agreement and estimating the coefficient for a single measure (ICC(A,1) model), as defined by McGraw & Wong (1996), was employed. While there are several different versions of the ICC and some confusion regarding model choice (Weir, 2005), two-way mixed models with an absolute agreement definition are suggested to be the appropriate form of ICC to apply when the goal is to assess test-retest reliability (Koo & Li, 2016). Specifically, this decision was affected by the fact that it was expected to observe test-retest effects (i.e., the observation of slightly differing results upon the second measure as compared to the first measure) due to factors such as the process of familiarizing with the setup and procedure during the two test runs. That is, an absolute agreement definition was chosen as the presence of systematic error (i.e., bias) was expected, and thus to which extent the repeated measures agreed was of interest. A paired t-test was performed in order to assess the presence of systematic error between the two test runs. Furthermore, participants were randomly assigned, whereas the measures were determined by the fixed research question, constituting a mixed design. As for rater-effects, these were considered absent, seeing that raters were blind to the syllable presentation during response collection.

Considering that the number of trials has a substantial impact on the reliability of the experiment (Pedhazur & Schmelkin, 2013), reliability was calculated for both the full-length paradigm including 120 trials, as well as for the first 40 (i.e., blocks one vs. two) and 80 trials (i.e., blocks one and two vs. blocks three and four). This was done to evaluate the test-length effect of the present paradigm. Reliability values may range from 0 (no reliability) to 1 (perfect reliability), with values closer to 1 indicating stronger reliability. Interpretation of r_{ICC} estimates was based on suggestions reported by Koo and Li (2016, p. 161) in which "values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability". As further recommended, interpretations additionally took the estimated confidence intervals into account as the obtained ICC is only an expected value of the true ICC. More specifically, if the 95% confidence interval of an estimated value ranged from, for instance, 0.76 to 0.91, reliability was concluded as being "good" to "excellent."

Comparability of the manual- and the verbal-response format was evaluated (a) by intraclass correlations between the runs of the two versions, and (b) by using a mixed-model analysis of variance (ANOVA). The ANOVA was set up as a three-factorial design, including the repeated-measure factors *Repetition* (the first vs. the second run of the paradigm) and *Response Format* (manual vs. verbal), as well as the between-subject factor *Sex*. The dependent variable was the LI. Statistical analyses were conducted in IBM SPSS Statistics (version 25; IBM Corp., Armonk, NY), and GPower (version 3.1; www.gpower.hhu.de) was used for test power calculation. Effect sizes were expressed as Cohen's *d* or proportion explained variance (η^2).

2.5.1. Data availability statement. The data associated with the present results, as well as the dichotic task (E-Prime file; Psychology Software Tools, Pittsburgh, PA), are openly available and can be downloaded for free from the OSF platform (https://osf.io/aj26n/).

3. Results

The mean laterality of the manual-response format was $M_{\mathrm{LI}} = 29.94$ ($SD_{LI} = 22.72$) at first testing and $M_{\mathrm{LI}} = 34.37$ ($SD_{LI} = 22.82$) upon retest. For the verbal paradigm, the mean was $M_{\mathrm{LI}} = 33.21$ ($SD_{LI} = 22.92$) at the first run of the paradigm and $M_{\mathrm{LI}} = 37.42$ ($SD_{LI} = 22.30$) for the retest. All four LI were significantly larger than 0, all $t(49) \geq 9.32$, all $p < .001$, all Cohen's $d > 1.27$. Mean differences between the two runs of the test were $M_{diff} = 4.21$ ($SD_{diff} = 8.61$) for the manual test and $M_{diff} = 4.42$ ($SD_{diff} = 7.30$) for the verbal, both $t(49) > 3.50$, all $p < .001$, all $d > 0.5$. Mean values for the correct reports of left- and right-ear stimuli are presented in Table 3.

The analysis of the full paradigm (i.e., 120 trials) yielded a reliability of $r_{ICC} = .93$, $p < .001$, 95% CI [.82, .97] for the manual-, and $r_{ICC} = .91$, $p < .001$, 95% CI [.82, .96], for the verbal-response paradigm, respectively. Considering a single block of 40 trials in the manual paradigm, the reliability was $r_{ICC} = .64$, 95% CI [.25, .82]. Considering two blocks (i.e., 80 trials) it was $r_{ICC} = .90$, 95% CI [.82, .94]. For a single block in the verbal-response version, the reliability was $r_{ICC} = .65$, 95% CI [.08, .85]. For two blocks, $r_{ICC} = .86$, 95% CI [.76, .92]. Reliability estimates for correct reports of left- and right-ear stimuli are provided in Table 3.

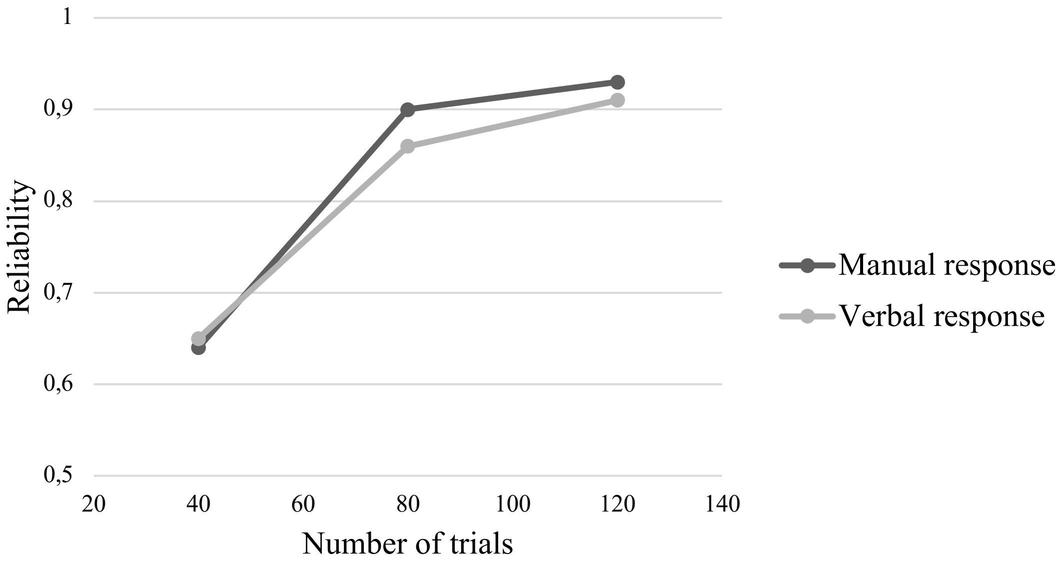

Figure 3. Reliability of the present paradigm as a function of the number of trials.

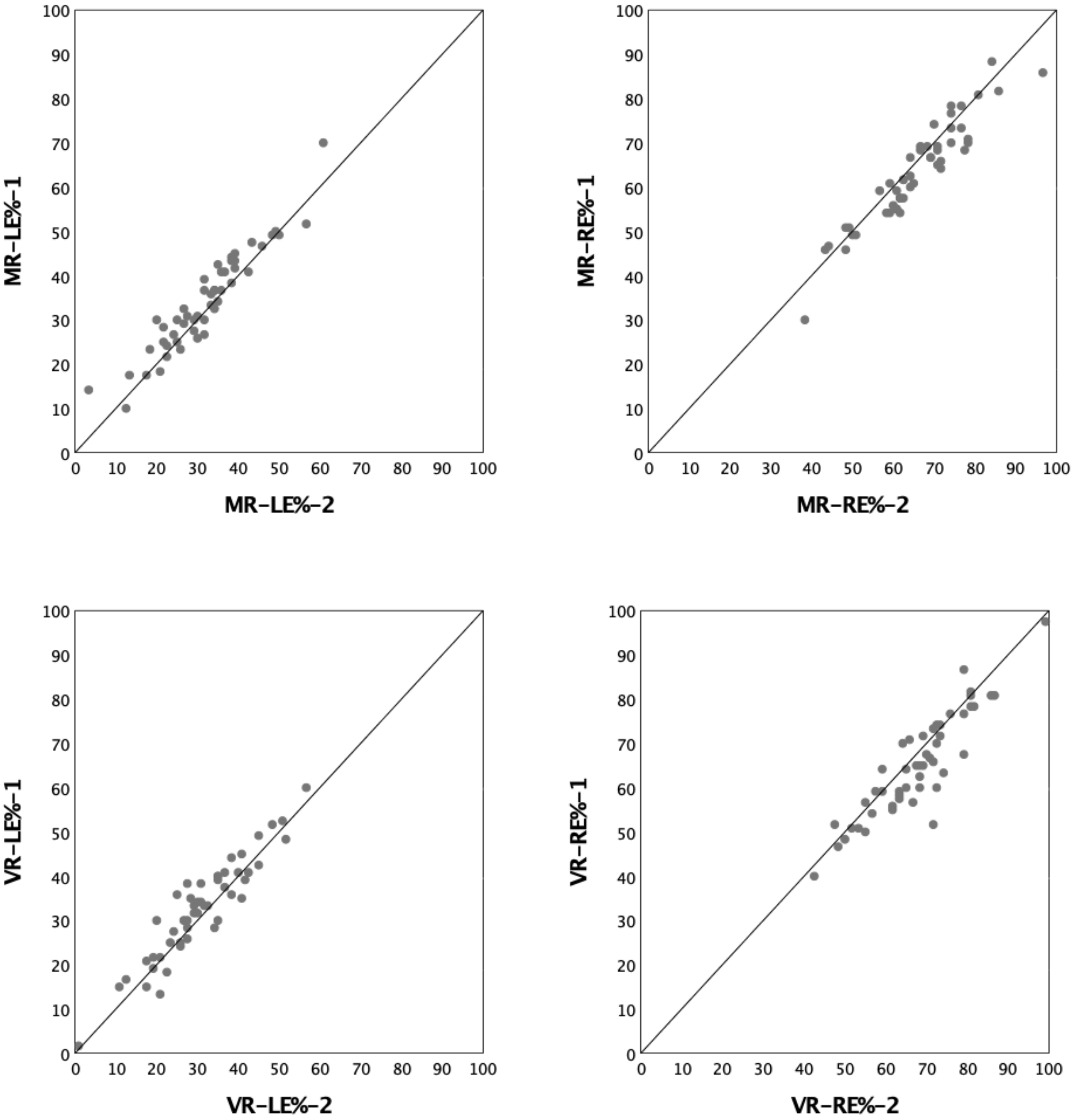

Figure 4. Scatterplots of left- and right-ear correct reports in test-retest of the manual- and verbal-response formats of the present paradigm, respectively. MR = manual-response format; VR = verbal-response format; LE% = percentage of correct left-ear reports; RE% = percentage of correct right-ear reports. Numbers 1 and 2 refer to the first run (i.e., test) and the second run (i.e., retest) of the paradigm.

Table 3

Mean values, standard deviation (SD), and reliability (r_{ICC})[a] of correct recall of left- (L_c) and right-ear stimuli (R_c)

		Left ear			**Right ear**		
Response format		*M*	*SD*	r_{ICC} (95% CI)	*M*	*SD*	r_{ICC} (95% CI)
Manual	Test	40.92	13.47	.93 (.82, .97)	75.94	13.73	.93 (.82, .97)
	Retest	38.32	13.31		78.62	13.92	
Verbal	Test	39.10	13.53	.92 (.85, .96)	78.06	13.88	.88 (.73, .94)
	Retest	37.02	13.19		81.36	13.44	

Note. (a) Intraclass correlations determined as ICC(A,1).

In comparing the manual- and verbal-response versions, an r_{ICC} = .71, 95% CI [.54, .82], was found of the LI of the first manual test with the LI of the first verbal test, and an r_{ICC} = .75, 95% CI [.53, .86], with the second verbal test. For the second run of the manual test, the correlations were r_{ICC} = .77, 95% CI [.63, .86] with the first verbal test, and r_{ICC} = .83, 95% CI [.72, .90] with the second verbal test run.

A significant main effect of *Repetition* [$F(1, 48) = 30.71$, $p < .009$] was found in the ANOVA, although with small effect size [$\eta^2 = 0.01$], suggesting a stronger LI at retest as compared to first testing across both response formats. The main effect of *Response Format* was not significant [$F(1, 48) = 2.28$, $p = .14$, $\eta^2 < 0.01$]. A sensitivity power analysis (at 5% alpha probability, and with a $r = .80$ correlation of the repeated measures) suggested that medium to large effects (> 3% variance explained) could be excluded with a power of .80. Neither the interaction of *Repetition* and *Response Format* [$F(1, 48) = 0.04$, $p = .84$, $\eta^2 < 0.01$], nor any of the main or interaction effects including the factor *Sex* [all $F(1, 48) < 1$, all $p > .50$, all $\eta^2 < 0.01$] were significant.

4. Discussion

The aim of the present study was to design and evaluate a novel dichotic-listening paradigm optimized for the reliable assessment of hemispheric specialization for speech processing. In line with the hypothesis, the here-suggested paradigm, irrespective of response format, reveals "good" to "excellent" test-retest reliability. With estimates having a 95% chance of landing within the range of .82 to .97, the full paradigm yielded results of the utmost standard as compared with the typically reported values for dichotic-listening paradigms (cf. Speaks, Niccum, & Carney, 1982; Hugdahl & Hammar, 1997; Voyer, 1998; Gadea et al., 2000; Voyer & Rodgers, 2002; Hugdahl, 2011; Westerhausen, 2019). That is, previous retest correlations of paradigms with the same length (i.e., 120 trials) have reported values between a minimum of $r = .63$ (Wexler & King, 1990; using rhyming words as stimuli) and a maximum of $r = .91$ (Wexler, Halwes & Heninger, 1981; using VCV-stimuli). However, even when shortened to 80 trials, the present paradigm reaches "moderate" to "good" reliability with estimates in-between .58 and .86. This suggests that it might not be necessary to include as many as 120 trials to achieve reliable estimates when using the present design features. Indeed, even when utilizing as many as 90 or 100 trials, previous studies have tended to yield retest correlations of values lesser than the lower bounds of the 95% confidence intervals in the present 80-trial paradigm (Voyer, 1998). For instance, with the use of CV-syllables and the instruction of reporting back both stimuli presented within a trial, a retest reliability of $r = .71$ was reported in a 90-trial paradigm (Speaks et al., 1982). Then again, the range of the confidence intervals for the present 80-trial version is wider than for the full-length paradigm, suggesting that the 120-trial version of the paradigm should be preferred. Irrespective of the number of trials, however, the here-suggested paradigm—in comparison with already established paradigms—appears to indicate reliability estimates in the upper end of the dichotic-listening field. It thus appears that the design features implemented here certainly improves reliability for dichotic-listening testing.

From test to retest, correct recalls increased slightly for the right ear, whereas there was a slight decrease for the left ear in both the manual and the verbal paradigm. Accordingly, the overall difference in average LI between the two test runs was 4.2% (± 8.6%) and 4.4% (± 7.3%) in the manual and the verbal version, respectively. Thus, considering that changes were confirmed to be present upon the repeated measure, it appears that either test-retest effects such as initial coding errors may have occurred, or that the differences could be attributed to inadequate design features such as an insufficient attempt of syllable fusion. Nevertheless, seeing that these changes were relatively small, it seems unreasonable to believe that their emergence had much to do with the design features of the present paradigm. Rather, it is believed that test-retest effects were accountable for most of the differences between the two runs, as was expected prior to testing. The retest-reliability for ear scores were "good" to "excellent" for both the left and the right ear in the manual version, indicating that the correct recalls were consistent between the first test and retest. In the verbal-response paradigm, however, "good" to "excellent" estimates were exclusive to the left ear. The right ear scores in the verbal test revealing "moderate" to "excellent" estimates appear to confirm that test-retest effects were present, as well as indicating that the verbal response might have been a less appropriate format than the manual response. The effects of response mode will be discussed later. Nevertheless, in line with the well-established phenomenon discovered by Kimura (1961a, 1961b), it was made clear that the majority of the subjects revealed a consistent REA during both test and retest, and within both response formats (overall LI = 33.7%).

By calculating r_{ICC} values for 40, 80, and 120 trials within the present paradigm, it was demonstrated that reliability was positively correlated with the number of trials (see Figure 3). According to the Spearman-Brown formula for test-length adjustment, the number of trials is unquestionably a major factor affecting reliability (Pedhazur & Schmelkin, 2013). Intuitively, it might seem that a further increase in the number of trials would yield even higher estimates of reliability. In 1996, Marc Brysbaert even argued that any reliability estimate could be reached by using a sufficient number of trials (Voyer, 1998). However, reliability will eventually see diminishing returns with the increase of trials. By applying the Spearman-Brown prediction formula to the present paradigm, an increase from 120 to 160 or 200 trials can be predicted to only marginally increase the reliability. For instance, the reliability of the verbal paradigm would be predicted to increase from .93 to .95 and .96, respectively. The indication for such diminishing returns within both the verbal and the manual versions of the present paradigm can also be observed in Figure 3. Previous studies using more than 120 trials have also failed to report considerable improvements and even observed reduced estimates for higher numbers of trials (Hiscock, Cole, Benthall, Carlson, & Ricketts, 2000; Speaks et al., 1981). For instance, Hiscock et al. (2000) reported that an increase from 120 to 480 trials would only yield a slight reliability increase from .85 to .96, according to the Spearman-Brown formula. Additionally, it should come as no surprise that extending the test length carries the risk of tiring the participant as well as the experimenter, increasing the chances of random errors – especially in clinical samples. Taken together, administering a test length of 120 trials appears to be an entirely feasible yet highly reliable solution.

It should be addressed that the above comparison of paradigm reliabilities has to account for the fact that most previous studies have reported reliability estimates as product-moment correlations (e.g., Wexler et al., 1981; Speaks et al., 1982; Wexler & Halwes, 1983; Wexler & King, 1990; Hugdahl & Hammar, 1997; Voyer, 1998; Gadea et al., 2000; Hiscock et al., 2000; Voyer & Rodgers, 2002). At least two short-comings appear to emerge with the use of product-moment correlations in test-retest studies (McGraw & Wong, 1996). Firstly, they do not consider the fact that the variables share a common measurement class. More specifically, repeated measures (consisting of multiple variables sharing a common measurement class) are used in retest studies, whereas product-moment correlations are intended to use when measuring the relationship between variables representative of different measurement classes (e.g., the relation of caffeine consumption to sleep quality). Secondly, the mean differences between the variables are ignored with product-moment correlations as the calculation of the correlation coefficient only accounts for the covariance of the variables. Put another way, as opposed to the ICC(A,1), product-moment correlations are not sensitive to systematic error (Weir, 2005). With these issues in mind, the here reported intraclass correlations appear to represent more appropriate estimates of retest reliability than does common reports of product-moment correlations in previous dichotic-listening studies (Koo & Li, 2016). To illustrate, the product-moment correlation (Pearson's r) of the manual and verbal paradigm would be $r = .95$ ($p < .001$) and $r = .93$ ($p < .001$), respectively, as opposed to the estimated $r_{ICC} = .93$ and $r_{ICC} = .91$ in the present study. While it is obvious that Pearson's r reveals higher estimates than the intraclass correlation coefficient of the present data, the question of what it actually represents—whether it refers to the *degree* or *direction* of laterality—is yet to be answered. Above all, product-moment correlations as used in measuring the retest reliability of laterality measures may not only lead to spurious correlations but are also difficult to interpret as they appear to make incorrect assumptions about the underlying distribution of the data (McManus, 1983; Voyer, 1998; Gadea et al., 2000). From an arguable perspective, the common misapplication of product-moment correlations in assessing the test-retest reliability of dichotic-listening paradigms might owe to the idea that "when hundreds of thousands of papers written by eminent researchers have been published using the same statistical approach, this approach becomes a kind of law against which no one thinks to go" (Berchtold, 2016, p. 6). For this reason, when comparing the present paradigm with previous paradigms, it should be recognized that previous studies have had a tendency to overestimate the "true" reliability of their paradigms.

Correlating the measures of both response formats (manual and verbal) revealed values within the range of $r_{ICC} = .53$ to $r_{ICC} = .90$. The wide range of these values, which additionally contains a lower bound well below the estimates of the full paradigm ($r_{ICC} = .53$ vs. $r_{ICC} = .82$), appears to indicate that the two paradigms measure slightly different aspects of laterality. During the verbal response, for instance, the essential process of speech production might modulate the perceptual laterality (Westerhausen, 2019). Interestingly, depending on whether speech comprehension or speech production is chosen as the consideration of interest, the Wada-test (sodium amobarbital), as used for assessing validity, has previously yielded different results (Voyer, 1998). Additionally, the experimenter also has to register and fully understand the oral response, and any problems occurring on this end might, eventually, lead to coding errors. In the present data, for instance, it was observed that several participants had reported syllables that were not by any means present in a given trial. Thus, given that this reoccurring event was discovered in both response formats, a likely explanation is that these reports were due to logging errors made by either the participant or the experimenter. Alternatively, regarding manual response, the participant is required to adapt to a novel response system, which might be considered a challenging task by itself. While younger generations might display an advantage in the process of learning computerized response systems, it could potentially become especially challenging for digitally illiterate populations such as older adults to adapt to such a system (e.g., Mead et al., 2000). Whereas this should be kept in mind when testing such populations, the present sample did not indicate any notable disadvantage at this stage. As for all age groups, however, familiarizing with and handling the manual-response setup and procedure requires additional response mapping and selection processes, including visual-motor coordination (Van den Noort, Specht, Rimol, Ersland, &

Hugdahl, 2008). Although it could be argued that additional familiarization trials could be implemented for attenuating trial-to-trial variability (Weir, 2005), the feasibility of the experiment would conceivably suffer from such an implementation. Nonetheless, with no observers (i.e., experimenters) involved in the reporting process, the manual format might be considered less prone to error, given that there is no room for interindividual logging inaccuracies. At the same time, the main effect of Response Format was non-significant and small, whereas the test power was sufficient to exclude considerable mean LI differences between the two versions with some confidence. Nevertheless, the reduced correlations appear to suggest that small differences between the manual- and verbal-response modes exist. Low ICC values, however, may imply either a lack of agreement between the two measures, lack of sample variability, or an insufficient sample size (Koo & Li, 2016). Given the elsewhere high ICC estimates yielded from the same sample, it appears that the sample size did not account for these reduced values. Hence, it might be that these reduced correlations refer to either a reduction in agreement or that the sample variability was reduced between the formats. Thus, it remains currently not possible to state whether one of the two versions is a better option for the measurement of hemispheric specialization.

Regarding sex, there was not found any significant main effect nor interaction effects on the LI. Laterality differences in sex have previously been reported as significant, though they are found to be small (Voyer, 1998; Voyer, 2011; Hirnstein, Westerhausen, Korsnes, & Hugdahl, 2013). The fact that sex differences in lateralization were not observed in the present study might indicate either that (1) previously detected differences have occurred as a result of suboptimal design features, or (2) differences exist, but the present sample was not sufficient to observe them. Indeed, non-significant differences between the sexes do not necessarily imply that these differences do not exist (Rich-Edwards, Kaiser, Chen, Manson, & Goldstein, 2018). If point (1) is true, it appears that sex differences have mistakenly occurred as a product of measurement imprecision. More precisely, sex differences might be attributable to differences in the uncontrolled, top-down effects induced by the task (Voyer, 2011). Within the present paradigm, factors introducing error variance has been controlled for to a greater degree than others have done before, and it would thus be expected to see that many of the previously reported effects in dichotic listening could be ruled out. However, if point (2) turns out to be true, small sex differences might exist at the population level, as previously concluded by Voyer (2011). Notably, factors that have previously been reported to affect sex differences in dichotic listening, including age (Hirnstein et al., 2013) and menstrual cycle (e.g., Hodgetts, Weis, & Hausmann, 2015), might have been under-represented within the present sample. Nevertheless, the measures obtained here are representative of a rather limited sample as compared with the above-cited studies and is therefore not sufficient to add to the controversy over sex differences in dichotic-listening performance. The sample was, however, sufficient for the present purpose (Koo & Li, 2016).

In evaluating the retest reliability of the current paradigm, a sample with an age range from 19 to 39 years was used. A plethora of studies have been conducted to investigate the associations between age and cognition, many of which have had important implications for the planning and execution of dichotic-listening studies. For instance, it has been demonstrated that an inverted U-shaped association of age and cognitive-control performance can be found within speech perception (Westerhausen, Bless, Passow, Kompus, & Hugdahl, 2015). That is, both children and elderly individuals have been found to display reduced cognitive-control performance in the forced-left condition of the forced-attention paradigm. Young adults, on the other hand, have demonstrated an overall better performance in detecting stimuli. Correspondingly, it appears that the REA remains stable throughout the young and middle adulthood as opposed to an increased magnitude in older age (Westerhausen et al., 2015). By addressing such findings, the present study decided to exclusively include young and middle-aged adults for investigation. However, considering that the need for cognitive control has been significantly reduced within the present paradigm, it can be speculated on whether an expanded age range, including both younger and older populations, would reveal similar estimates. Alternatively, it would be fruitful to establish whether adjustments compensating for disadvantages associated with the different age groups could be made to the testing procedure. For instance, a recent study on an aging sample avoided the exclusion of individuals with hearing deficits by adjusting the dB SPL of the stimulus presentation, compensating for individual differences in terms of hearing threshold (Passow et al., 2012). Thus, to conclude whether the here-obtained measures are generalizable to other populations, the present paradigm needs to be further investigated in the future.

By controlling for the known reliability-modulating components and improving the stimulus presentation, it became obvious that the reliability increased as compared with previous paradigms. Intuitively, it appears that the attempt to reduce the relevance of higher cognitive functions was successful. These results also seem to extend the “Hugdahl et al.” model by revealing that an initial bottom-up processing indeed exists. Then again, it can only be speculated with caution on whether the yielded estimates reflect “purer” laterality indices than other paradigms. In an extensive review by Westerhausen & Hugdahl (2008), for instance, it was suggested that the corpus callosum plays a major role in the top-down processing of dichotic stimuli. More specifically, it was proposed that an initial, predisposed bottom-up processing fosters the REA, whereas an additional top-down processing enables an inter-hemispheric transfer, leaving the ear advantage prone to modulation. If this is true, there is reason to believe that the here-reduced relevance of top-down processes supports the initial bottom-up processing and, subsequently, allows for less demand on the corpus callosum. Nevertheless, the exploration of such theories is beyond the scope of this study. As for now, interpretations of the neural substrate underlying the present findings remain speculative.

4.1. Implications

Previous dichotic-listening paradigms appear to not only have revealed suboptimal reliability estimates, but also to have overestimated their reliability indexes by using inappropriate calculations. With reportedly more appropriate formulas for quantifying test-retest reliability (McGraw & Wong, 1996; Koo & Li, 2016), and subsequently reliability estimates in the upper end of the dichotic-listening field, the here-suggested paradigm may provide important implications to both research and diagnostics. As for experimental settings, many of the previously observed effects and developed theories within dichotic listening could be confirmed or ruled out as the reliability of the present paradigm appear to be indicative of optimized design features. For instance, heavily debated topics such as the source of the ear advantage (see Hiscock & Kinsbourne, 2011; Pollmann, 2010) and sex differences (e.g., Voyer, 2011) in dichotic-listening performance could benefit from new insights revealed with the application of a more reliable paradigm. However, it should be noted that the present design is aimed towards the exclusive study of stimulus-driven laterality in speech processing. Nevertheless, several implications derived from the present paper, such as the critical choice of reliability index, considerations of response format, and the effects of reducing the relevance of higher-order cognition are all highly relevant topics for most researchers dealing with the planning and execution of cognitive experiments. For instance, no previous studies have thoroughly considered the effects of response format on dichotic-listening performance (Westerhausen, 2019). Thus, this might be considered a significant contribution to the current literature – at least a basis for further investigation.

With respect to clinical and developmental populations, the paradigm introduces a task that may be considered rather effortless to perform, with presumably no prerequisites for cognitive-control capacity required. Most importantly, however, is the fact that reliable measures are essential to establish before applying any method in the clinic (Koo & Li, 2016).

Seen that dichotic paradigms of mediocre reliability have been utilized within clinical settings (e.g., the forced-attention paradigm; for reliability, see Hugdahl & Hammar, 1997), the present paradigm thus appears to offer a better-suited alternative. Until now, the forced-attention paradigm has been excessively used to assess atypical asymmetries in disorders such as schizophrenia (Hiscock & Kinsbourne, 2011). It has, for instance, been concluded that patients with schizophrenia struggle to perform top-down modulations (Hugdahl, 2003; Hugdahl et al., 2009). Assuming that the interpretation of the present results is valid, by applying the present approach, for example, one could aim to assess the bottom-up laterality of schizophrenia patients to evaluate whether the observed deficit is exclusive to top-down modulation. In this way, the here-presented paradigm might also serve to support previously drawn conclusions. Finally, while there is a growing popularity of neuroimaging techniques, and even paradigms attempting to combine dichotic listening with fMRI emerge (Van den Noort et al., 2008), the present paradigm provides a rather cheap and feasible alternative.

4.2. Limitations

There are limitations in the present study that should be pointed out. Right-handed adult university students and employees with good hearing acuity and high identification rates of homonyms comprised the sample in the present study. Thus, it remains to be established whether data similar to the here presented reliability estimates can be reproduced outside these inclusion criteria, especially in clinical and aging populations. Then again, there was also room for even fairer testing conditions between participants. The sample could have consisted exclusively of native speakers and, even more specifically, spoken dialects could have been taken into account (see Future Directions). Due to a restricted sample, the results were not sufficient to, for instance, confirm the existence of sex differences. Although this was not a major focus in the present study, it would be interesting to see how the paradigm would perform in larger samples. Operating with two different labs also had its limitations in that lacking the exact same equipment has been unavoidable. With the collection of additional data, such as reaction times, the results could have been provided with possibly a more detailed understanding of the source of the errors. Although the present study has controlled for the recognized variables modulating the reliability of verbal dichotic-listening paradigms, other variables are yet to be identified. It should indeed be acknowledged that the existence of other reliability-modulating variables than the ones covered here is not at all denied. As for the ICC, which appeared to be of appropriate use, it also has to be acknowledged that, as an index of relative reliability and not absolute reliability (Weir, 2005), the evaluation could have benefited from combining the ICC with additional measures. Regarding validity, the here utilized stimulus material has proven to be valid (Westerhausen, 2019). However, the validity of the current paradigm also needs to be confirmed - a task that was not currently possible to perform. That is, due to the lack of validation, it is not presently possible to state with full confidence that the paradigm functions as a tool for investigating structural laterality.

4.3. Future Directions

Although the stimulus material was deemed appropriate by all participants, it may be suspected that certain syllables sounded more familiar to specific native speakers. It has been argued that CV-syllables are semantically meaningless (Hugdahl et al., 2009). However, the syllable /ka/ is, for instance, a highly frequent word of use in Norway (meaning “what” in several Norwegian spoken dialects). Thus, it might be speculated that, within trials consisting of a given syllable paired with /ka/, the participants speaking these dialects favored /ka/ as a result. There is, nevertheless, little to no evidence on the effects of word frequency and neighborhood density within dialects on laterality measures. Moreover, the syllables /ba/, /da, /ta/ and /ga/ act as meaningful and frequently used words in the Norwegian language. As seen earlier, however, the stimulus material of the CV is strictly limited to six syllables only, and it would have been especially challenging to match the stimuli for equal item difficulty. Nevertheless, whether spoken dialects affect ear preference would be an interesting study. Unfortunately, differences in spoken dialects of the subjects were not registered as part of the present study. It was, therefore, difficult to examine whether any of the measures were affected by this. If possible, it is thus encouraged to account for dialects in future research to explore these effects. It is believed that the application of the current paradigm to populations speaking the same dialects would make for even fairer testing conditions between participants and, ultimately, yield higher estimates within an experimental test-retest study.

4.4. Concluding Remarks

Through an evaluation of the here-suggested paradigm, it became evident that its design features substantially improved dichotic-listening reliability. A shorter version of the present paradigm with no more than 80 trials is applicable, although 120 trials appear to be ideal for highly reliable testing. Both a verbal- and a manual-response version of the paradigm is available. At the moment, both response versions appear to be equally suitable. Here, no sex differences were present on the LI. It is speculated that either previously reported sex differences in dichotic listening could have occurred as a result of inadequate testing or that the sample of the present study was not sufficient to detect any existing differences. Although the age groups tested consisted of young and middle-aged adults, there is also the speculation that the paradigm will apply to younger and older populations considering that the need for exerting cognitive effort has been significantly reduced. While previous studies have often reported suboptimal estimates, they have, at the same time, tended to overestimate the "true" reliability by using less appropriate correlation coefficients for evaluating reliability. This should be taken into account when comparing the present paradigm with previous paradigms.

Dichotic listening is often interpreted as a two-stage process consisting of an initial bottom-up stage and a second stage attributing additional cognitive processes (Hiscock & Kinsbourne, 2011; Hugdahl et al., 2009; Westerhausen et al., 2013). In the present study, the task at hand was to minimize cognitive demands in the attempt to assess, within this context, the early auditory, perceptual first stage in a reliable manner. This is a challenging task considering the complex mechanisms of the human brain and the vast number of variables known to modulate perception. Furthermore, in evaluating tools for such assessment, Pedhazur & Schmelkin (2013) has also warned that test-retest reliability should be used with caution as low estimates may suggest either unreliable measures, changes in the measured individual, or both. However, it is believed that the here-yielded estimates points in the direction towards an optimized dichotic-listening paradigm for the assessment of hemispheric dominance for speech processing. The present findings are thus interpreted as indicating that within free-report dichotic-listening paradigms examining the perceptual laterality of speech processing, retest reliability serves as a function of the cognitive demands presented by the task and its conditions. That is, as soon as the requirement for involving executive functions increases, the obtained measures are prone to become less reliable - and vice versa.

In summary, a highly reliable dichotic-listening paradigm has been designed. This was done by (a) reducing the cognitive demands of the task and (b) improving essential features of the stimulus presentation. Point (a) is also accompanied by the beneficial side effect of making individual- and group-differences in higher cognitive functions less likely to affect the obtained measures, again allowing for fair testing in both clinical and developmental samples. Point (b) minimizes the number of trials required for balanced design and, consequently, for yielding reliable measures, supporting the opportunity of conducting economic assessments of laterality - at least in experimental settings.

References

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1-29.

Badzakova-Trajkov, G., Häberling, I. S., Roberts, R. P., & Corballis, M. C. (2010). Cerebral asymmetries: Complementary and independent processes. *PloS one, 5*(3), e9682.

Berchtold, A. (2016). Test-retest: Agreement or reliability? *Methodological Innovations, 9*, 1-7

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 106*(18), 7351-7356.

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological Review, 64*(3), 205-215.

Bryden, M. (1988). An overview of the dichotic listening procedure and its relation to cerebral organization. In K. Hugdahl (Ed.), *Handbook of dichotic listening: Theory, methods and research* (pp. 1-43). Chichester: Wiley & Sons.

Bryden, M., Munhall, K., & Allard, F. (1983). Attentional biases and the right-ear effect in dichotic listening. *Brain and Language, 18*(2), 236-248.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45-50.

Carey. D. P., & Johnstone, L. T. (2014). Quantifying cerebral asymmetries for language in dextrals and adextrals with random-effects meta analysis. *Frontiers in Psychology, 5,* 1128.

D’Anselmo, A., Marzoli, D., & Brancucci, A. (2016). The influence of memory and attention on the ear advantage in dichotic listening. *Hearing Research, 342,* 144-149.

Dos Santos Sequeira, S., Specht, K., Moosmann, M., Westerhausen, R., & Hugdahl, K. (2010). The effects of background noise on dichotic listening to consonant-vowel syllables: an fMRI study. *Laterality, 15,* 577-596.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23,* 83-94.

Gadea, M., Gomez, C., & Espert, R. (2000). Test-Retest Performance for the Consonant-Vowel Dichotic Listening Test With and Without Attentional Manipulations. *Journal of Clinical and Experimental Neuropsychology, 22*(6), 793-803.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129-135.

Hellige, J. B. (1993). *Hemispheric asymmetry: What's right and what's left*. Cambridge, MA: Harvard University Press.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402.

Hirnstein, M., Westerhausen, R., Korsnes, M. S., & Hugdahl, K. (2013). Sex differences in language asymmetry are age-dependent and small: A large-scale, consonant-vowel dichotic listening study with behavioral and fMRI data. *Cortex, 49*(7), 1910-1921.

Hiscock, M., Cole, L. C., Benthall, J. G., Carlson, V. L, & Ricketts, J. M. (2000). Toward solving the inferential problem in laterality research: Effects of increased reliability on the validity of the dichotic listening right-ear advantage. *Journal of the International Neuropsychological Society, 6,* 539-547.

Hiscock, M., Hampson, E., Wong, S. C. P., & Kinsbourne, M. (1985). Effects of eye movements on the recognition and localization of dichotic stimuli. *Brain and Cognition, 4,* 140-155.

Hiscock, M., Inch, R., & Ewing, C. (2005). Constant and variable aspects of the dichotic listening right-ear advantage: A comparison of standard and signal detection tasks. *Laterality: Asymmetries of Body, Brain and Cognition, 10*(6), 517-534.

Hiscock, M., & Kinsbourne, M. (2011). Attention and the right-ear advantage: What is the connection? *Brain and Cognition, 76*(2), 263-275.

Hodgetts, S., Weis, S., & Hausmann, M. (2015). Sex hormones affect language lateralization but not cognitive control in normally cycling women. *Hormones and Behavior, 74*, 194-200.

Hugdahl, K. (1991). Brain lateralization: Dichotic listening studies. In B. Smith & G. Adelman (Eds.), *Neuroscience Year 2: Encyclopedia of neuroscience* (pp. 23-26). Boston: Birkhauser.

Hugdahl, K. (2003). Dichotic Listening: An Experimental Tool in Clinical Neuropsychology. In K. Hugdahl (Ed.), *Experimental Methods in Neuropsychology* (pp. 29-46). Norwell, MA: Kluwer Academic Publishers.

Hugdahl, K. (2011). Fifty years of dichotic listening research – Still going and going and. *Brain and Cognition, 76*, 211-213.

Hugdahl, K. (2016). Dichotic listening and attention: the legacy of Phil Bryden. *Laterality: Asymmetries of Body, Brain and Cognition, 21*, 1-22.

Hugdahl, K., & Andersson, L. (1986). The “forced-attention paradigm” in dichotic listening to CV-syllables: A comparison between adults and children. *Cortex*, *22*, 417-432.

Hugdahl, K., & Hammar, Å. (1997). Test-Retest Reliability for the Consonant-Vowel Syllables Dichotic Listening Paradigm. *Journal of Clinical Experimental Neuropsychology, 19*(5), 667-675.

Hugdahl, K., Westerhausen, R., Alho, K., Medvedev, S., & Hämäläinen, H. (2009). Attention and cognitive control: Unfolding the dichotic listening story. *Scandinavian Journal of Psychology, 50*(1), 11-22.

Kimura, D. (1961a). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology/Revue Canadienne de psychologie, 15*(3), 166-171.

Kimura, D. (1961b). Some effects of temporal-lobe damage on auditory perception. *Canadian Journal of Psychology/Revue Canadienne de psychologie, 15*(3), 156-165.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex, 3,* 163-178.

Kimura, D. (2011). From ear to brain. *Brain & Cognition, 76*(2), 215-217.

Kinsbourne, M. (1970). The cerebral basis of lateral asymmetries in attention. *Acta Psychologica, 33,* 193-201.

Kompus, K., Specht, K., Ersland, L., Juvodden., H. T., van Wageningen, H., Hugdahl, K., & Westerhausen, R. (2012). A forced-attention dichotic listening fMRI study on 113 subjects. *Brain and Language, 121*(3), 240-247.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155-163.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and Hearing, 19*(1), 1-36.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods, 1*(1), 30-46.

McManus, I. C. (1983). The interpretation of laterality. *Cortex, 19,* 187-214.

Mead, S. E., Sit, R. A., Rogers, W. A., Jamieson, B. A., & Rousseau, G. K. (2000). Influences of general computer experience and age on library database search performance. *Behaviour & Information Technology, 19*(2), 107-123.

Musiek, F. E., & Weihing, J. (2011). Perspectives on dichotic listening and the corpus callosum. *Brain and Cognition, 76*, 225-232.

Ocklenburg, S. (2017). Tachistoscopic Viewing and Dichotic Listening. In L. J. Rogers & G. Vallortigara (Eds.), *Lateralized Brain Functions* (pp. 3-28). New York: Springer Verlag.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Passow, S., Westerhausen, R., Wartenburger, I., Hugdahl, K., Heekeren, H. R., Lindenberger, U., & Li, S. C. (2012). Human Aging Compromises Attentional Control of Auditory Perception. *Psychology and Aging*, *27*(1), 99-105.

Pedhazur, E. J., & Schmelkin, L. P. (2013). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Penner, I. K., Schläfli, K., Opwis, K., & Hugdahl, K. (2009). The role of working memory in dichotic-listening studies of auditory laterality. *Journal of Clinical and Experimental Neuropsychology, 31*(8), 959-966.

Plack, C. J. (2004). Auditory perception. In C. J. Plack (Ed.), *The Sense of Hearing* (pp. 1-19). Abingdon, Oxon: Psychology Press.

Pollmann, S. (2010). A unified structural-attentional framework for dichotic listening. In K. Hugdahl & R. Westerhausen (Eds.), *The two halves of the brain: Information processing in the cerebral hemispheres* (pp. 441-468). Cambridge, MA: MIT Press.

Repp, B. H. (1977). Measuring laterality effects in dichotic listening. *The Journal of the Acoustical Society of America, 62*(3), 720-737.

Rich-Edwards, J. W., Kaiser, U. B., Chen, G. L., Manson, J. E., & Goldstein, J. M. (2018). Sex and Gender Differences Research Design for Basic, Clinical, and Population Studies: Essentials for Investigators. *Endocrine Reviews, 39*(4), 424-439.

Rimol, L. M., Eichele, T., & Hugdahl, K. (2006). The effect of voice-onset-time on dichotic listening with consonant-vowel syllables. *Neuropsychologia, 44*(2), 191-196.

Rogers, L. J. (2014). Asymmetry of brain and behavior in animals: Its development, function, and human relevance. *Genesis, 52*(6), 555-571.

Speaks, C., Blecha, M., & Schilling, M. (1980). Contributions of monotic intelligibility to dichotic performance. *Ear and Hearing, 1*(5), 259-266.

Speaks, C., Niccum, N., & Carney, E. (1982). Statistical properties of responses to dichotic listening with CV nonsense syllables. *The Journal of the Acoustical Society of America, 72*(4), 1185-1194.

Speaks, C., Niccum, N., Carney, E., & Johnson, C. (1981). Stimulus dominance in dichotic listening. *Journal of Speech, Language, and Hearing Research, 24*(3), 430-437.

Strouse Carter, A., & Wilson, H. R. (2001). Lexical effects on dichotic word recognition in young and elderly listeners. *Journal of the American Academy of Audiology, 12*(2), 86-100.

Studdert-Kennedy, D., Shankweiler, D., & Schulman, S. (1970). Opposed Effects of a Delayed Channel on Perception of Dichotically and Monotically Presented CV Syllables. *The Journal of the Acoustical Society of America, 48*(2B), 599-602.

Sætrevik, B., & Hugdahl, K. (2007a). Endogenous and exogenous control of attention in dichotic listening. *Neuropsychology, 21*(3), 285-290.

Sætrevik, B., & Hugdahl, K. (2007b). Priming inhibits the right ear advantage in dichotic listening: Implications for auditory laterality. *Neuropsychologia, 45*(2), 282-287.

Techentin, C., & Voyer, D. (2010). Word frequency, familiarity, and laterality effects in a dichotic listening task. *Laterality, 16*(3), 313-332.

Tremblay, P., Deschamps, I., Baroni, M., & Hasson, U. (2016). Neural sensitivity to syllable frequency and mutual information in speech perception and production. *NeuroImage, 136*, 106-121.

Van den Noort, M., Specht, K., Rimol, L. M., Ersland, L., & Hugdahl, K. (2008). A new verbal reports fMRI dichotic listening paradigm for studies of hemispheric asymmetry. *Neuroimage, 40*(2), 902-911.

Voyer, D. (1998). On the Reliability and Validity of Noninvasive Laterality Measures. *Brain and Cognition, 36*, 209-236.

Voyer, D. (2011). Sex differences in dichotic listening. *Brain and Cognition*, 76, 245-255.

Voyer, D., & Rodgers, M. A. (2002). Reliability of laterality effects in a dichotic listening task with nonverbal material. *Brain and Cognition, 48*(2-3), 602-606.

Voyer, D., & Techentin, C. (2009). Dichotic listening with consonant-vowel pairs: The role of place of articulation and stimulus dominance. *Journal of Phonetics, 37*(2), 162-172.

Voyer, D., Dempsey, D., & Harding, J. A. (2014). Response procedure, memory, and dichotic emotion recognition. *Brain and Cognition, 85,* 180-190.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research, 19*, 231-240.

Wester, K., Lunde, A. J., Hugdahl, K., & Taksdal, I. (1998). Dichotic memory: Paradoxical effect of removing a left frontal gyrus: A case study. *International Journal of Neuroscience, 93*(3-4), 279-86.

Westerhausen, R. (2019). A primer on dichotic listening as a paradigm for the assessment of hemispheric asymmetry. *Laterality: Asymmetries of Body, Brain and Cognition, 24*(6), 740-771.

Westerhausen, R., Bless, J. J., & Kompus, K. (2015). Behavioral laterality and aging: The free-recall dichotic-listening right-ear advantage increases with age. *Developmental Neuropsychology, 40*(5), 313-327.

Westerhausen, R., Bless, J. J., Passow, S., Kompus, K., & Hugdahl, K. (2015). Cognitive control of speech perception across the lifespan: A large-scale cross-sectional dichotic listening study. *Developmental Psychology*, *51*(6), 806-815.

Westerhausen, R., & Hugdahl, K. (2008). The corpus callosum in dichotic listening studies of hemispheric asymmetry: A review of clinical and experimental evidence. *Neuroscience and Biobehavioral Reviews, 32*, 1044-1054.

Westerhausen, R., & Kompus, K. (2018). How to get a left-ear advantage: A technical review of assessing brain asymmetry with dichotic listening. *Scandinavian Journal of Psychology, 59*, 66-73.

Westerhausen, R., Moosmann, M., Alho, K., Medvedev, S., Hämäläinen, H., & Hugdahl, K. (2009). Top-down and bottom-up interaction: manipulating the dichotic listening ear advantage. *Brain Research, 1250*, 183-189.

Westerhausen, R., Passow, S., & Kompus, K. (2013). Reactive cognitive-control processes in free-report consonant-vowel dichotic listening. *Brain and Cognition, 83*(3)*,* 288-296.

Wexler, B. E., & Halwes, T. (1983). Increasing the power of dichotic methods: The fused rhymed words test. *Neuropsychologia*, *21*(1), 59-66.

Wexler, B. E., Halwes, T., & Heninger, G. R. (1981). Use of a statistical significance criterion in drawing inferences about hemispheric dominance for language function from dichotic listening data. *Brain and Language, 13*, 13-18.

Wexler, B. E., & King, G. P. (1990). Within-modal and cross-modal consistency in the direction and magnitude of perceptual asymmetry. *Neuropsychologia, 28*, 71-80.

Zaidel, E. (1983). Disconnection syndrome as a model for laterality effects in the normal brain. In J. B. Hellige (Ed.), *Cerebral Hemisphere Asymmetry: Method, Theory, and Application* (pp. 95-151). New York: Praeger.

Appendix A

Volunteers Wanted!

Test yourself in an experiment on cognition.

- Duration: about 1 hour
- Location: Institutt for biologisk og medisinsk psykologi (IBMP), 9th floor, BB-Bygget, Jonas Lies vei 91
- Compensation: a gift voucher (150 kr)
- Only pre-condition: right-handedness

Questions? Interested? Please get in touch via email: ihiexperiment@gmail.com

The data collection is part of the project: "*Inter-hemispheric integration in working memory*", coordinator: rene.westerhausen@psykologi.uio.no

Experiment on cognition
ihiexperiment@gmail.com

Experiment on cognition
ihiexperiment@gmail.com

Experiment on cognition
ihiexperiment@gmail.com

Experiment on cognition
ihiexperiment@gmail.com

Experiment on cognition
ihiexperiment@gmail.com

Experiment on cognition
ihiexperiment@gmail.com

Experiment on cognition
ihiexperiment@gmail.com

Experiment on cognition
ihiexperiment@gmail.com

Experiment on cognition
ihiexperiment@gmail.com

Experiment on cognition
ihiexperiment@gmail.com

Experiment on cognition
ihiexperiment@gmail.com

Appendix B

Participant Information Sheet

Dear participant,

Thank you for considering participating in our research project on auditory attention. The *purpose* of our project is to study how the two halves of the brain communicate with each other during speech perception. We will study this by presenting you speech stimuli via headphones and showing visual stimuli on the screen. The study recruits healthy, right-handed participants with normal hearing and without history of psychiatric/neurologic disease. Please read this information carefully before deciding whether to participate in the study or not.

The Task. It is a computerized experiment which allows us to examine attention processes in speech perception. In this task you will hear verbal stimuli, which are presented via headphones. These stimuli are syllables, like "ba" and "ka", spoken by a male voice. The stimuli are presented two at the same time, that is, one sound will be presented to your left ear, and a different one to your right ear. The experiment will be divided into 2 "blocks". Each block will last about 20 min. You will get to practice this task for a few minutes before your performance is recorded. There will be short breaks between blocks.

Confidentiality and Data protection. All study data collected during the experiment will be stored, analysed, and published in anonymous form, that is, it will not be possible to link the data to your person. Anonymization will take place immediately after the end of the experiment. Your contact details will *not* be stored, and it will not be possible to relate the collected experiment data to your person. All staff is subject to professional secrecy and data protection compliance. Should you decide to participate, you will need to sign a Participant Consent Form (PCF), which will be kept in a locked cupboard with no reference to the experiment data.

Withdrawal of Consent. Participation in the study is voluntary. You have the right to withdraw your consent to participate in the study at any time without providing any reason, and without penalty. You also can demand us to delete your data at any time, which, however, will only be possible before anonymization of the data.

Thank you for your interest and support! If you have any questions, please ask the experimenter.

René Westerhausen (Professor, Department of Psychology, University of Oslo)
e-mail: rene.westerhausen@psykologi.uio.no, tel.: +47 22845230

Participation Consent Form

Study: *Inter-hemispheric integration in auditory working memory (IHI-WM)*

please tick boxes
for confirmation
⬇

☐ I confirm that I have read and understood the participant information sheet for the above study (document: IHI-aWM Version 1.1; dated 05 Jan. 2018). I have had the opportunity to consider the information, ask questions, and have had these answered satisfactorily.

☐ I understand that my *personal information* will not be stored (except for this Participant Consent From).

☐ I understand that all *experiment data* will be stored and analysed in an anonymised form, and it will not be possible to relate the study data to my person

☐ I understand that my participation is voluntary and that I am *free to withdraw at any time*, without giving any reason, without penalty.

☐ I understand that I can demand the collected data to be deleted at any time before the anonymization

☐ I understand that this project has been reviewed by and received ethics clearance through the Department of Psychology's Research Ethics Committee, University of Oslo (Ref.: 3838804).

☐ I agree to take part in the above study.

Name of Volunteer	Place, Date	Signature
Experimenter	Place, Date	Signature

Appendix C

ID:

Sex: [ ] male; [ ] female

Age:

HANDEDNESS INVENTORY

Please indicate your preferences in the use of hands in the following activities **by putting** **“+”** in the appropriate column. Where the preference is so strong that you would never try to use the other hand unless absolutely forced to, **put** “++”. If in any case you are really indifferent put **“+“ in both** columns.

Some of the activities require both hands. In these cases, the part of the task, or object, for which hand preference is wanted is indicated in brackets.

		LEFT	RIGHT
1	**Writing**		
2	**Drawing**		
3	**Throwing**		
4	**Scissors**		
5	**Toothbrush**		
6	**Knife** (without fork)		
7	**Spoon**		
8	**Broom** (upper hand)		
9	**Striking Match** (match)		
10	**Open box** (lid)		