Applicability and outcomes of a test for binaural phase sensitivity in elderly listeners

INGA HOLUBE[1,3], THERESA NUESSE[1,3], OLAF STRELCYK[2], ANNAEUS WILTFANG[1], PETRA VON GABLENZ[1,3] AND ANNE SCHLUETER[1,3]

[1] *Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, D-26121 Oldenburg, Germany*

[2] *Sonova U.S. Corporate Services, Cincinnati, OH, USA*

[3] *Cluster of Excellence "Hearing4All", Oldenburg, Germany*

Interaural phase difference (IPD) discrimination in the binaural auditory system has been shown to be related to localization abilities and speech intelligibility in background noise. One of the tests for binaural phase sensitivity determines the highest frequency of the test tone for which an interaural phase difference of 180° is detectable (IPD-FR). This test was included in a test battery together with examination of visual and hearing abilities, balance, tactile- and motor-skills, and cognitive abilities. The IPD-FR test was conducted with an adaptive 3-AFC experiment starting with a test-tone frequency of 250 Hz. Sixty-five of 220 participants could not perform the IPD-FR task. The main predictors for inability to perform the IPD-FR task are hearing loss at 250 Hz, fluid intelligence, measurement number, and gender. A linear regression analysis revealed that the test result IPD-FR threshold is related to pure-tone thresholds at low frequencies, composite score of cognition, and composite score of tactile sensitivity, fine motor skills, and vision, as well as gender. A correlation analysis shows that the IPD-FR threshold is not related to speech recognition in non-dynamic listening conditions with speech from the front if low-frequency hearing loss is taken into account.

INTRODUCTION

Speech recognition performance appears to be influenced by multiple factors such as hearing loss, age, cognitive abilities, and supra-threshold auditory processing. One of the factors for supra-threshold auditory processing is the sensitivity to interaural phase differences (IPD) in the binaural auditory system. IPD has been shown to be related to localization abilities and speech intelligibility in background noise (see e.g., Strelcyk and Dau, 2009). Several test paradigms are used in audiological research to measure IPD. One of the tests determines the highest frequency for which an IPD of 180° is detectable (IPD-FR threshold, test names "IPD-FR" in Neher *et al.*, 2011, and "TFS-AF" in Füllgrabe *et al.*, 2017). The performance of participants in these kinds of tests can be affected by personal characteristics such as age and hearing loss (see Füllgrabe and Moore, 2018 for a meta-analysis). Furthermore, Füllgrabe and Moore

*Corresponding author: inga.holube@jade-hs.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 7: Auditory Learning in Biological and Artificial Systems, August 2019, Nyborg, Denmark. Edited by A. Kressner, J. Regev, J. C.-Dalsgaard, L. Tranebjærg, S. Santurette, and T. Dau. The Danavox Jubilee Foundation, 2019. ISSN: 2596-5522.

(2018) assumed that cognitive abilities might have an influence on test performance as a substantial amount of variance could not be explained by other factors.

The IPD-FR test was regarded as a test for supra-threshold auditory processing ability. It was included in an extensive test battery containing audiological standard diagnostics, questionnaires, motor and vision skills as well as cognitive tasks to analyze their relation to speech recognition. As a relatively large number of elderly listeners was unable to perform the IPD-FR test, we investigated: (i) Prediction of the individual inability to perform the IPD-FR task, (ii) Explanation of variance in the measured IPD thresholds.

METHODS

Participants

Two hundred and twenty-three volunteers (77 from Hörzentrum Oldenburg GmbH database and 146 from public announcement) aged from 55 to 81 years participated in an extensive test battery (see Table 1). Three participants were excluded due to single sided deafness or technical issues. All participants were numbered based on the order of their participation date (measurement number).

	Age cohorts (years)					
	55-60	**61-65**	**66-70**	**71-75**	**76-81**	**all**
Number of participants	51	43	46	42	38	220
Number female/male	33/18	21/22	25/21	22/20	20/18	121/99
Median PTA-4 (dB HL)	12.5	20.6	27.5	35.6	34.4	23.8
Median PTA-low (dB HL)	9.4	15.6	15.3	20.6	20.6	14.4
IPD-FR ability/inability	48/3	31/12	31/15	23/19	22/16	155/65
Mean IPD-FR threshold (Hz)	757.4	678.0	657.5	574.0	549.8	664.9

Table 1: Characteristics of the participants. PTA-4 denotes the average hearing loss at 0.5, 1, 2, and 4 kHz. PTA-low denotes the average hearing loss at 0.25, 0.5, 0.75, and 1 kHz averaged between both ears.

IPD-FR

An adaptive 3-AFC experiment (1-up-2-down rule) was used to determine the highest frequency for which an IPD of 180° was detectable (Neher *et al.*, 2011). The outcome of the experiment was the IPD-FR threshold in Hz. Within each interval, a sinusoid with a duration of 2 s and an amplitude modulation of 1 Hz was presented over headphones. The two reference intervals included diotic stimuli whereas the interaural phase changed between 0 and 180° every 0.5 s in the target interval. The presentation level of all stimuli was 30 dB SL. The experiment started with a frequency of 250 Hz and a step size of 250 Hz. The step size was halved after each upper of a total of eight reversals. A minimum bound of 125 Hz and a maximum of 20 kHz was chosen for the IPD-FR threshold. Oral instructions were given to the participants and a training with 10 trials at minimum was carried out. All participants except the first 29 administered the task using a touch screen.

Test battery

Pure-tone hearing thresholds were measured via air conduction (0.125 to 8 kHz) and via bone conduction (0.5 to 6 kHz).

Several tests were used to determine cognitive skills, visual abilities, fine motor skills, and tactile sensitivity. To ensure audibility of instructions during those tests, participants wore hearing aids if German indication criteria (G-BA, 2017) were fulfilled. Hearing aids (Phonak Bolero V90-P or V90-SP depending on the severity of hearing loss) were fitted to those participants who did not own hearing aids. Hearing aid owners could decide, based on their subjective preference, whether they wanted to use their own or the newly fitted hearing aids.

1) Composite scores of cognitive tests:

The test outcomes of the cognitive tests were z-transformed and averaged to obtain three composite scores related to a typical categorization of cognitive abilities:

- Fluid intelligence: Ruff2 & 7 (Ruff and Allen, 1996), Trail-Making Test A and B (Reitan, 1992), STROOP (Puhr and Wagner, 2012), TAP divided attention test (Zimmermann and Fimm, 2013), digit span forward and backward (Petermann, 2012)
- Crystallized intelligence: Regensburg word fluency test (Aschenbrenner *et al.*, 2000), multiple choice vocabulary test (Lehrl, 2005)
- Verbal memory: Verbal learning and memory test (Helmstaedter *et al.*, 2001)

2) Vision, fine motor skills and tactile sensitivity

Three outcomes for visual abilities, fine motor skills, and tactile sensitivity were used as single values. These skills might be necessary to handle the touch screen in the IPD-FR task although the keys on the screen were large and easy to catch.

- Tactile sensitivity was measured as the 75%-threshold of spatial resolution for fingertips using JVP domes (Johnson *et al.*, 1997).
- Fine motor skills: The MLS test battery of *Schuhfried* (Neuwirth and Benesch, 2012) was applied. The outcomes were standardized and combined to four factors (factors 1-3, and 5 in Neuwirth and Benesch, 2012). For further analysis, only factor 5 characterizing the movement speed of arm, hand, and fingers was used.

- Visual acuity was measured bespectacled if glasses were prescribed and available using *Optovist* (VISTEC Vision Technologies, 2010).

3) Speech recognition

Speech levels for recognition scores of 50% (SRT) for speech in quiet were determined with the German Freiburg digit test and the Göttingen sentence test (GÖSA; Kollmeier and Wesselkamp, 1997). The SRT for speech in background noise was measured for GÖSA in the standardized stationary Gönoise using headphones and via loudspeaker in the TASCAR system (Grimm *et al.*, 2015). In two other TASCAR conditions, IFFM (Holube *et al.*, 2010) and a cafeteria recording were used as noise sources. The target sentences were always presented from the front. Gönoise and IFFM were also presented from the front and the cafeteria recording from all eight loudspeakers of the system. In all speech tests, the participants wore the same hearing aids (Phonak Bolero V90-P/SP, see above) fitted according to NAL-NL2 fitting procedure (Keidser *et al.*, 2011) if they fulfilled the German indication criteria. In the speech tests, in contradiction to the other tests described in (1) and (2), the participants were not allowed to use their own hearing aids if available to ensure similar amplification schemes in the speech tests.

RESULTS

Predicting individual inability to perform the IPD-FR task

Sixty-five participants (approx. 30%) could not detect the IPD of 180° at any frequency down to 125 Hz. The age distribution of the inability groups relative to the ability group is additionally given in Table 1. Logistic regressions were used to find predictors of inability to perform the IPD-FR task. In the following model calculations, 5 participants were excluded due to outliers in the cognitive variables.

In the first step of the model development process, air conduction pure-tone threshold at 250 Hz was identified as having the highest significance of all audiogram parameters. This variable was kept in the model and further single variables were added (but not kept). Significant contributions were found for fluid intelligence ($p < 0.001$), measurement number ($p = 0.004$), age ($p = 0.010$), touch screen usage ($p < 0.012$), and crystallized intelligence ($p = 0.037$). There was no gender effect ($p = 0.057$). Verbal memory and none of the vision acuity, tactile, or motor variables were significant.

In a next step, the significant variables and female gender were added to the logistic model starting with fluid intelligence and kept if they had a significant contribution to the model. Otherwise, they were omitted. In addition, a combination (mean) of fluid and crystallized intelligence was tested, but resulted in a lower R^2 compared to fluid intelligence alone. The procedure resulted in a model with four independent variables shown in Table 2 ($R^2_{(\mathrm{Nagelkerkes})} = 0.576$, sensitivity = 90.7%, specificity = 65.6%).

The odds (likelihood) for inability increases by a factor of 1.78 per 5 dB increase in hearing loss in the worse ear at 250 Hz. Besides this, fluid intelligence, but not crystallized intelligence or verbal memory, has a significant contribution to the model. The order of the participants, coded in the measurement number, contributes significantly to inability. The odds for inability increases e.g., by a factor of 1.1 per decreasing of measurement number by 10. The measurement number includes: A) Non-usage of a touch screen for the first 29 participants. B) Recruitment of the first 77 participants from the database of Hörzentrum Oldenburg GmbH. On average, those participants had a higher hearing loss and included more hearing aid users compared to those from public announcement. Therefore, this variable might include hearing loss differences not covered by the threshold at 250 Hz. C) Possible training effects including optimization of the oral instructions of the examiner. D) Other unknown temporal or recruiting effects. In addition, gender has a small but significant effect: The odds for inability is 2.8 times higher in female than in male.

Independent variables	**Wald**	***p***	**Odds ratio (95% C.I.)**
AC WEHL 250 Hz	38.7	< 0.001	1.12 (1.08-1.16)
Fluid intelligence	12.0	0.001	4.07 (1.84-8.99)
Measurement number	7.4	0.007	1.01 (1.00-1.02)
Female gender	5.4	0.020	2.79 (1.18-6.60)

Table 2: Logistic regression model for inability of the IPD-FR task. AC WEHL 250 Hz: air conduction thresholds in the worse ear at 250 Hz.

Predicting IPD-FR thresholds

A linear regression analysis was calculated for the 145 participants who could perform the IPD-FR task and for whom complete composite scores of cognitive skills and tactile/motor/vision were available. Dropouts were equally distributed over age groups.

In a first step, PTA-low was identified as most significant variable of all audiogram variables in the linear regression model. Thereafter, each variable was included in the linear regression model (but not kept). Measurement number, touch screen usage, and verbal memory did not contribute significantly to the model. The other variables were added stepwise to the linear regression model and kept if they had a significant contribution. To reduce the number of variables, several of them were combined: A significant correlation between the cognitive composite z-scores for fluid and crystallized intelligence of 0.499 ($p < 0.001$) was observed. Therefore, the mean of both cognitive z-scores was used as a combined variable. In addition, fine motor skills, tactile sensibility, and vision acuity were significantly correlated (motor-tactile: 0.210, $p = 0.011$; motor-vision: 0.181, $p = 0.029$; tactile-vision: 0.244, $p = 0.003$).

Hence, the z-scores of all three variables were averaged forming a new variable tactile/motor/vision. The resulting linear regression model with four independent variables is shown in Table 3.

Independent variables	R	R^2 corr.	R^2 change	Coeff B
PTA-low	0.371	0.132	0.132	-7.0
Mean of fluid and crystalline intelligence z-score	0.468	0.208	0.076	101.9
Tactile/Motor/Vision z-score	0.532	0.268	0.060	95.0
Male gender	0.584	0.322	0.054	113.4

Table 3: Results of the linear regression model with four independent variables.

Table 3 reveals that the IPD-FR threshold decreased by about 70 Hz for an increase in mean low-frequency pure-tone threshold by 10 dB. The IPD-FR threshold increased by about 100 Hz for an increase in cognitive and tactile/motor/vision skills by one standard deviation. Although, it has to be pointed out that the two independent variables are significantly correlated ($r = 0.359$, $p < 0.001$). In addition, males had on average about 110 Hz higher IPD-FR thresholds than females. It should be noted that similar to the logistic regression, age had no additional significant contribution to the model when the cognitive abilities were taken into account.

RELATION OF IPD-FR THRESHOLD TO SPEECH RECOGNITION

IPD-FR threshold was significantly correlated with the SRT for all speech test conditions measured with hearing aids. Spearman's correlation coefficients are 0.240 to 0.362 ($p \leq 0.004$). The highest correlation coefficient was observed for the fluctuating masker IFFM. When controlling for low-frequency pure-tone thresholds, the absolute values of partial correlations between IPD-FR threshold and SRTs dropped to the range from 0.015 to 0.146 and were no longer significant ($0.082 \leq p \leq 0.851$).

DISCUSSION

The first objective of this study was to predict the individual inability to perform the IPD-FR task using logistic regression. Personal characteristics that predicted the inability were hearing loss in the worse ear at 250 Hz, fluid intelligence, measurement number, and gender. Especially the effect of the factor measurement number left some open questions. Generally, it was surprising that the IPD-FR task was an insuperable obstacle for many participants. In contradiction to this observation, Füllgrabe *et al.*

(2018) stated that “reliable threshold estimates can be obtained relatively quickly […] and without practice” for (nearly) all listeners. 14 out of the 65 participants who could not perform the task met the inclusion criteria of Füllgrabe *et al.* (2018), i.e. air-conduction thresholds up to 1.5 kHz ≤ 25 dB HL. For future studies, more training (e.g., with ILD cues) might improve the percentage of participants who can perform the test and increase specificity. Furthermore, the IPD-FR minimum of 125 Hz might be too high (compared to 30 Hz in TFS-AF) and the start frequency should be lowered if participants were not able to use the cue at 250 Hz.

In a second step, the IPD-FR thresholds of those participants who were able to perform the task were analyzed with the aim to explain the variance. As presumed by Füllgrabe and Moore (2018) in their meta-analysis and consistent with Strelcyk *et al.* (2019), cognitive abilities were significantly predictive. An additional influence of the parameter “age” on the thresholds was not found in the present data. However, other sensory and motor skills also had predictive power.

With regard to the lack of correlation between IPD-FR threshold and speech recognition when controlling for low-frequency hearing loss, the applied speech tasks should be reconsidered. The relationship might be observable only for more spatial and dynamic speech conditions (e.g., Neher *et al.*, 2011).

ACKNOWLEDGEMENT

Thanks to Olga Schwarz and Anika Morgenstern for conducting part of the measurements, as well as Ulrike Lemke, Sigrid Scherpiet, and Emily Urry for project support. Research funded by the governmental funding initiative “Niedersächsisches Vorab” of the Lower Saxony Minstry for Science and Culture in cooperation with Sonova AG (Stäfa, Switzerland).

REFERENCES

Aschenbrenner, S., Tucha, O., and Lange, K.W. (**2000**). “RWT: Regensburger Wortflüssigkeits-Test,” Göttingen Hogrefe.

Füllgrabe, C., Harland, A.J., Sek, A.P., and Moore, B.C.J. (**2017**). “Development of a method for determining sensitivity to temporal fine structure,” Int. J. Audiol., **56**, 926-935. doi: 10.1080/14992027.2017.1366078.

Füllgrabe, C., and Moore, B.C.J. (**2018**). “The association between the processing of binaural temporal-fine-structure information and audiometric threshold and age: A meta-analysis,” Trends Hear., **22**. doi: 10.1177/2331216518797259.

Füllgrabe, C., Sek, A.P., and Moore, B.C.J. (**2018**). “Senescent changes in sensitivity to binaural temporal fine structure,” Trends Hear., **22**, 1-16. doi: 10.1177/2331216518788224.

Grimm, G., Luberadzka, J., Herzke, T., and Hohmann, V. (**2015**). “Toolbox for acoustic scene creation and rendering (TASCAR): Render methods and research applications,” In: Proceedings of the Linux Audio Conference, Johannes Gutenberg University (JGU). Mainz, Germany.

Helmstaedter, C., Lendt, M., and Lux, S. (**2001**). "VLMT: Verbaler Lern- und Merkfähigkeitstest," Göttingen, Hogrefe.

Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (**2010**). „Development and analysis of an International Speech Test Signal (ISTS)," Int. J. Audiol., **49**(12), 891–903. doi: 10.3109/14992027.2010.506889.

Johnson, K.O., van Boven, R.W., and Phillips, J.R. (**1997**). "J.V.P. Domes for cutaneous spatial resolution measurement. Operation Manual," Stoelting Co., Wood Dale, Illinois (USA).

Keidser, G., Dillon, H., Flax, M., Ching, T.Y.C., and Brewer, S. (**2011**). "The NAL-NL2 prescription procedure," Audiol. Res., **1**(e24), 88-90. doi: 10.4081/audiores.2011.e24.

Kollmeier, B., and Wesselkamp, M. (**1997**). "Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment," J. Acoust. Soc. Am., **102** (4), 2412–2421. doi: 10.1121/1.419624.

Lehrl, S. (**2005**). "Mehrfachwahl-Wortschatz-Intelligenztest: Manual mit Block MWT-B," 5th edition, Balingen Spitta Verlag.

Neher, T., Laugesen, S., Søgaard Jensen, N., and Kragelund, L. (**2011**). "Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities?" J. Acoust. Soc. Am., **130**, 1542-1558. doi: 10.1121/1.3608122.

Neuwirth, W., and Benesch, M. (**2012**). "Manual Motorische Leistungsserie. Kurzbezeichnung MLS," Mödling: Schuhfried.

VISTEC Vision Technologies (**2010**). "Optovist. Gebrauchsanweisung ab Softwareversion 1.0.007," VISTEC AG - Olching (Germany).

Petermann, F. (Hg.) (**2012**). "Wechsler-Adult-Intelligence-Scale-Fourth Edition (WAIS-IV). Manual," Frankfurt/Main Pearson Assessment and Information GmbH.

Puhr, U., and Wagner, M. (**2012**). "Handanweisung Interferenztest nach Stroop. Kurzbezeichnung STROOP," 25th edition, Mödling SCHUHFRIED GmbH.

Reitan, R. M. (**1992**). "Trail Making Test: Manual for administration and scoring," Tucson Reitan Neuropsychology Laboratory.

Ruff, RM., and Allen, C.C. (**1996**). "Ruff 2 & 7 Selective Attention Test: Professional Manual," Lutz PAR. Retrieved from www.parinc.com.

Strelcyk, O., and Dau, T. (**2009**). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," J. Acoust. Soc. Am., **125**, 3328-3345. doi: 10.1121/1.3097469.

Strelcyk, O., Zahorik, P., Shehorn, J., Patro, C., and Derleth, R.P. (**2019**). "Sensitivity to interaural phase in older hearing-impaired listeners correlated with nonauditory trail making scores and with a spatial auditory task of unrelated peripheral origin," Trends Hear., **23**, 1-21. doi: 10.1177/2331216519864499.

Zimmermann, P., and Fimm, B. (**2013**). " Testbatterie zur Aufmerksamkeitsprüfung," Version 2.3. Teil 1. 2nd edition, Herzogenrath Psytest.