DISSERTATION

Modeling Binaural Perception Based on the Complex Correlation Coefficient

Von der Fakultät für Medizin und Gesundheitswissenschaften der
Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels eines

Doktors der Naturwissenschaften
(Dr. rer. nat.)

angenommene Dissertation von

Bernhard Eurich
geboren am 31. August 1991
in Memmingen

Gutachter: Prof. Dr. Mathias Dietz
Weiterer Gutachter: Prof. Dr. Dr. Birger Kollmeier
Tag der Disputation: 18.12.2023

Abstract

Binaural hearing, or the benefit from listening with two ears, contributes to spatial hearing and therefore helps to determine the position of a heard sound as well as to perceptually segregate competing sound sources. It therefore facilitates navigation, orientation, and communication in challenging acoustic situations. More than 5 % of the world's population – and the trend is increasing – require rehabilitation for their hearing loss in order to enjoy equal opportunities in society. A central component is the treatment with hearing aids. In addition, participants with normal hearing or mild hearing loss use smart headphones with similar functionality in situations where sound localization is required, such as road traffic. Therefore, there is great demand to enable users of hearing technology to benefit from binaural hearing. This requires a good understanding of binaural hearing. Such is manifested in models that reflect binaural perception and thus reproduce the essential characteristics. If such models are also computationally efficient, they can also be incorporated into hearing algorithms to adapt the processing strategy based on the sound quality predicted by the model. However, for decades, models of binaural hearing have been established that fall short in these requirements. Specifically, models have assumed (1) axonal delays of several milliseconds in the brainstem to encode interaural time differences, whereas mammalian physiology suggests shorter delays encoding interaural phase differences, and (2) a lower spectral and temporal resolution in binaural hearing than in monaural hearing, whereas more recent results suggest that both monaural and binaural hearing access the resolution provided by basilar membrane filtering. This thesis aims to overcome these contradictions by providing models that are consistent with behavioral physiological characteristics and at the same time computationally simple enough to be incorporated into hearing algorithms. Therefore, this thesis (1) supports and uses the complex correlation coefficient to be an efficient and comprehensive description of mammalian binaural processing, (2) it shows that interference across frequency and time can explain the apparently lower binaural resolution, and (3) provides a simple, real-time applicable and at the same time powerful model for sound quality assessment. The models provided in this thesis can contribute to a better understanding of binaural hearing and therefore to a better assessment of the sound quality of hearing algorithms.

Zusammenfassung

Binaurales Hören, oder das Profitieren vom Hören mit zwei Ohren, trägt maßgeblich zum räumlichen Hören bei, und damit sowohl zur Bestimmung der Einfallsrichtung von Schall, als auch zur perzeptiven Trennung von gleichzeitigen Schallereignissen. Dies erleichtert die Navigation, Orientierung und Kommunikation in komplexen akustischen Umgebungen. Über 5 % der Weltbevölkerung – Tendenz steigend – benötigen Rehabitilationsmaßnahmen für ihren Hörverlust, um gesellschaftliche Chancengleichheit zu erlangen. Die zentrale Rolle spielt dabei die Versorgung mit Hörgeräten. Zusätzlich werden Kopfhörer mit ähnlichem Funktionsumfang von Normalhörenden und leicht Schwerhörenden häufig in Situationen genutzt, in denen Außengeräusche lokalisiert werden müssen, wie beispielsweise im Straßenverkehr. Daher besteht ein hoher Bedarf, binaurales Hören auch bei Verwendung von Hörsystemen zu ermöglichen. Dies erfordert ein tiefgreifendes Verständnis des binauralen Hörens. Um dieses Verständnis zu festigen und zu quantifizieren, werden Modelle benötigt, welche die effektive binaurale Verarbeitung im Gehirn nachbilden und die Kerneigenschaften der Wahrnehmung reproduzieren können. Modelle die zugleich recheneffizient sind, können auch in Hörsysteme integriert werden, um die Verarbeitungsstrategie auf Basis der vom Modell geschätzten Klangqualität anzupassen. Allerdings erfüllen die traditionellen Modelle des binauralen Hörens diese Anforderungen nicht im nötigen Maße. Konkret wurde bisher angenommen, dass zur Bestimmung des Zeitunterschieds zwischen dem Schalleinfall am linkem und rechtem Ohr im Hirnstamm Laufzeitunterschiede von mehreren Millisekunden ausgewertet werden. Jedoch ist bei Säugetieren deutlich wahrscheinlicher, dass Neurone im Hirnstamm auf die Auswertung von Phasendifferenzen zwischen den Ohren abgestimmt sind, was eine Auswertung lediglich kurzer Laufzeitdifferenzen bedeutet. Außerdem wurde angenommen, dass die Frequenz- und Zeitauflösung beim binauralen Hören geringer ist als beim monauralen Hören. Neuere Studien zeigen jedoch, dass beim monauralen und binauralen Hören jeweils die volle Auflösung, die die Basilarmembran zulässt, zur Verfügung steht. Ziel dieser Arbeit ist, die genannten Widersprüche zu überwinden. Hierfür werden Modelle entwickelt, die sowohl mit Erkenntnissen aus der Psychoakustik als auch aus der Physiologie konform, gleichzeitig aber so schlicht aufgebaut sind, dass sie in Hörsysteme integriert werden können. Dies wird erzielt indem (1) der komplexe Kreuzkorrelationskoeffizient als effiziente und schlüssige Beschreibung der binauralen Wahrnehmung und Verarbeitung bei Säugetieren etabliert wird, (2) indem gezeigt wird, dass Interferenzen über Frequenz und Zeit die vermeintlich niedrigere binaurale Auflösung erklären können, und (3) indem ein schlichtes, echtzeitfähiges und zugleich robustes Modell zur Schätzung von Klangqualität entwickelt wird. Die hier vorgestellten Modelle tragen dadurch zu einem besseren Verständnis der binauralen Wahrnehmung und in der Folge zu einer verbesserten Ermittlung von Klangqualität bei.

List of Publications

Journal papers

published

Eurich, B., Encke, J., Ewert, S. D., and Dietz, M. (**2022**). Lower interaural coherence in off-signal bands impairs binaural detection. The Journal of the Acoustical Society of America **151**(6), 3927–3936, doi: `10.1121/10.0011673`.

Eurich, B., and Dietz, M. (**2023**). Fast binaural processing but sluggish masker representation reconfiguration. The Journal of the Acoustical Society of America **154**(3), 1862–1870, doi: `10.1121/10.0021072`.

submitted

Eurich, B., Ewert, S. D., Dietz, M., and Biberger, T. (**submitted**). A computationally efficient model for combined assessment of monaural and binaural audio quality , submitted to: Journal of the Audio Engineering Society.

Conference abstracts and proceedings

Eurich, B., Encke, J., Ewert, S. D., and Dietz, M. (**02-2021**). Modeling binaural masking release based on interaural phase distribution , Podium Presentation, 44th ARO Annual Midwinter Meeting.

Eurich, B., Encke, J., Ewert, S. D., and Dietz, M. (**02-2022**). The complex-valued correlation coefficients across frequency channels accounts for binaural detection , Poster Presentation, 45th ARO Annual Midwinter Meeting.

Eurich, B., Encke, J., Ewert, S. D., and Dietz, M. (**03-2022**). The complex-valued correlation coefficient across frequency channels accounts for binaural detection , Oral Presentation, DAGA 2022, Stuttgart.

Eurich, B., and Dietz, M. (**02-2023**). Binaural processing – fast, sluggish, or both? , Poster Presentation, 46th ARO Annual Midwinter Meeting, Orlando, USA.

Eurich, B., Biberger, T., Ewert, S., and Dietz, M. (**2023**). Towards a Computationally Efficient Model for Combined Assessment of Monaural and Binaural Audio Quality in *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, European Acoustics Association, Turin, Italy, pp. 299–302, doi: `10.61782/fa.2023.0575`.

Dietz, M., Eurich, B., Dietze, A., Klug, J., and Encke, J. (**10-2022**). A model of binaural interaction based on two correlators , 24th International Congress on Acoustics.

Dietz, M., Encke, J., Dietze, A., and Eurich, Bernhard Klug, J. (**03-2023**). A model of binaural interaction based on two correlators , DAGA 2023 - 49. Jahrestagung für Akustik.

Dietz, M., Eurich, B., Klug, J., and Encke, J. (**2023**). Towards a comprehensive model of binaural processing The Journal of the Acoustical Society of America **154**(4_supplement), A235, doi: `10.1121/10.0023388`.

Contents

Abstract iii

Zusammenfassung v

List of Publications vii

1 Introduction 1

- 1.1 Hearing and Society 1
- 1.2 Binaural Hearing 3
- 1.3 Insights through models: Interference across frequency and time unravels binaural modeling 5
- 1.4 Applying models: Assessing sound quality of hearing and reproduction technology 5

References 7

2 The model concepts 10

- 2.1 Concepts underlying this thesis 10
- 2.2 Interference concept suggested in this thesis 15

References 18

3 Lower interaural coherence in off-signal bands impairs binaural detection 22

3.1 Abstract 23
3.2 Introduction 23
3.3 Description of the Model 28
3.4 Predictions of Binaural-Detection Datasets 32
3.5 Discussion 36
3.6 Conclusion 40
3.7 Acknowledgments 41
3.8 Appendix 41

References 43

4 Fast binaural processing but sluggish masker representation reconfiguration 48

4.1 Abstract 49
4.2 Introduction 49
4.3 Simulations 51
4.4 Discussion 61
4.5 Conclusions 66
4.6 Acknowledgements 66

References 67

5 A Computationally Efficient Model for Combined Assessment of Monaural and Binaural Audio Quality) 71

5.1 Abstract 72
5.2 Introduction 72
5.3 Model Description 75
5.4 Databases 81
5.5 Results 84
5.6 Discussion 87
5.7 Conclusion 93
5.8 Acknowledgement 94

References 95

6 General discussion and conclusion 101

6.1 Spectral and temporal interference of IPD statistics 101

6.2 Demonstrating the viability of the two-channel code for effective binaural modeling 105

6.3 Impact of the findings 109

6.4 Implications for future research 111

6.5 Conclusion 112

References 113

7 Glossary 120

CHAPTER 1

Introduction

1.1 Hearing and Society

Unaddressed hearing loss can degrade communication, hinder childrens' speech and language development and affect mental health (World Health Organization, 2021). Next to social isolation, lonliness and stigma (World Health Organization, 2023), also education and employment can be affected, entailing a negative impact on individual opportunities, society, and economy (World Health Organization, 2023; Davis and Hoffman, 2019). As of 2023, more than 430 million people in the world, corresponding to more than 5 % of the world population, need rehabilitation for their hearing disability. World Health Organization predicts this number to increase to over 700 million people by 2050 (World Health Organization, 2023), associated with the growth of global population and the increasing life expectancy (Olusanya *et al.*, 2014; Davis and Hoffman, 2019). Hearing loss is the third leading cause of years lived with disability (YLD). For people older than 70 years it is even the leading cause of YLD (Wilson and Tucci, 2021). McCormack and Fortnum (2013) state hearing aids as "the primary clinical management intervention for people with hearing loss". Hearing aids have the potential to contribute significantly to manage that "immense global health concern"' (Wilson and Tucci, 2021). As one example, people with hearing loss have been associated with an increased risk of all-cause dementia (Jiang *et al.*, 2023; Griffiths *et al.*, 2020), which is a top-ten cause of death worldwide (World Health Organization, 2020). Among those using hearing aids, no increased risk has been found (Jiang *et al.*, 2023). Furthermore, hearing rehabilitation using hearing aids has been shown to have a positive impact on quality of life [e.g., Lotfi *et al.* (2009), Brodie *et al.* (2018), Ferguson *et al.* (2017)].

However, despite these benefits, the majority of people with hearing loss do not have hearing aids or do not use them (McCormack and Fortnum, 2013). Various reasons have been identified: Besides psycho-social factors, poor sound quality and lack of comfort, in combination with limited benefit, play major roles (McCormack and Fortnum, 2013; Abrams and Kihm, 2015; Vaisberg *et al.*, 2021; Bennett *et al.*, 2018). Also, inequalities in access and affordability of hearing devices contribute to an unmet need of hearing healthcare as high as 67. . .86 % (Committee on Accessible and Affordable Hearing Health Care for Adults *et al.*, 2016)[1]. At the same time, improvements in digital hearing technology have been achieved. These include algorithms for noise reduction, beamforming and binaural interaction. Features like wireless connectivity and introduction of artificial intelligence further improved usability and functionality (You *et al.*, 2020; Seol and Moon, 2022). Devices sold directly to the consumer like over-the-counter hearing aids, smart headphones and hearables address users with mild to moderate hearing loss (Seol and Moon, 2022). This complements prescription hearing aids that address those with moderate to severe hearing loss. With 20 % of the world population having mild-to-complete loss in the better hearing ear, which makes hearing loss an “invisible disability” (Wilson and Tucci, 2021), there is a large group of people that could potentially benefit from low-barrier, consumer-friendly hearing technology.

In order to improve hearing technology, research relies on models of auditory perception. Models are involved because they

---

[1] The research presented in this thesis mainly aims to improve hearing restoration by providing knowledge and algorithms. However, it is important to mention another crucial factor of managing the global burden of hearing loss: Approximately 90 % of people with moderate to profound hearing impairment live in low- to middle income countries (Davis and Hoffman, 2019), where infections as well as noise exposure are more prominent causes of hearing loss (World Health Organization, 2023). Therefore, a crucial challenge is prevention of hearing loss. This includes support for hearing healthcare in countries with low- to middle income. With the United Nations’ sustainable development goals (United Nations, 2015), the World Health Assembly’s resolution on the prevention of deafness and hearing loss (World Health Organization, 2017) and the Convention on the Rights of Persons with Disabilities, a policy framework for global action has been installed (Davis and Hoffman, 2019).

1. manifest and quantify researchers' understanding of hearing,
2. are the basis of the signal processing that operates in hearing technology, as a hearing aid aims to compeensate for the hearing loss by performing a function similar to that of intact hearing,
3. can speed up development by providing outcome predictions such as instrumental sound quality assessments to monitor distortions arising from signal processing, as hearing aids are useless if distortions exceed the benefit,
4. can contribute to accurate diagnostics.

This thesis directly addresses the aspects (1) to (3) with aspect (4) supported by the proposed models and insights. It incorporates a novel concept of binaural modeling – the complex correlation coefficient γ. It was developed with my contribution and is described in chapter 2. γ is used in all three models proposed in this thesis, which includes:

two contributions to the understanding of hearing: The same conceptual idea on the role of interference processes in binaural perception, described in section 2.2, is applied in the frequency domain and in the time domain in order to unify previously contracting models of binaural hearing. Both are standalone peer-reviewed publications that are shortly outlined in section 1.3 and reprinted in the chapters 3 and 4.

one contribution to model-based sound quality assessment: The γ-based model presented in chapter 3 is extended to a combined monaural and binaural measure for audio quality assessment (eMoBi-Q). Combining perceptual validity, performance and computational efficiency it is suited for evaluation and real-time control of algorithms in hearing technology. Section 1.4 shortly outlines the submitted manuscript which is reprinted in chapter 5.

1.2 Binaural Hearing

In the following, binaural hearing is introduced, as it is the subfield of hearing research addressed in this thesis. Subsequently, the sections 1.3 and 1.4 give an overview of the binaural research conducted.

Binaural hearing describes the ability to evaluate the differences in a sound arriving at the two ears. It allows listeners to determine the horizontal position of the sound. Spatial hearing relies on binaural cues and spectral cues arising from the geometry of the outer hear, head, and torso. Together, spatial hearing enables to turn the head towards an object without a-priori knowledge of their direction and without any further cues (Masterton *et al.*, 1969). Therefore, (spatial) hearing is beneficial for orientation and identifying hazards (Brown, 1994; Grothe and Pecka, 2014). The auditory image analysis ("what is the source of the sound?") is thereby complemented by the auditory location anaylsis ("Where is the source of the sound?") (Brown, 1994). Combined with vision, a detailed analysis of the surrounding has developed under evolutionary selective pressure (Brown, 1994; Heffner, 1997).

In addition, binaural hearing helps segregating competing sound sources (Shinn-Cunningham, 2005), allowing listeners to focus on a sound source of interest in the presence of distracting sources (Shinn-Cunningham *et al.*, 2017). A target sound is better detected in the presence of a masking sound when the two sounds differ in their interaural parameters, known as binaural unmasking [BU, Hirsh (1948); Culling and Lavandier (2021)]. This contributes significantly to spatial release from masking (Dieudonné and Francart, 2019; Culling and Lavandier, 2021), i.e. improved speech intelligibility as a consequence of spatial sepearation of target speaker and masking noise, where monaural and binaural unmasking effects are combined (Dieudonné and Francart, 2019; Bronkhorst and Plomp, 1988). Additionally, in noisy, reverberant listening conditions, binaural listening reduces listening effort compared to monaural listening (Rennies and Kidd, 2018). As everyday situations involve conversations in noisy, often reverberant environments, as well as traffic, binaural hearing facilitates communication and safe navigation.

Therefore it is desirable that hearing aid algorithms, mainly aiming to restore speech intelligibility, preserve and provide binaural cues (Marquardt *et al.*, 2015; Derleth *et al.*, 2021). Designing and evaluating hearing technology, such as hearing aids or wireless headphones with a hear-through mode, requires a good understanding of hearing, especially binaural hearing.

1.3 Insights through models: Interference across frequency and time unravels binaural modeling

Models that replicate important aspects of human hearing are important to manifest a good understanding of hearing. For models about binaural perception, there have been contradictions between the assumptions underlying traditional models and more recent physiologic and behavioral insights. In both the spectral and temporal domain, experimental results have been interpreted to indicate that binaural hearing operates on larger analysis windows than monaural hearing. In both domains, this has been in conflict with other experimental results that suggest the same analysis windows operating in binaural and monaural hearing. This thesis suggests one unifying concept and applies it to unify the contradictions in both domains. The concept is based on interference across frequency and time and is introduced in section 2.2. **Chapter 3** and **chapter 4** show how this interference can unify the conflicts between binaural analysis windows in the spectral and temporal domain, respectively.

1.4 Applying models: Assessing sound quality of hearing and reproduction technology

As mentioned above, improvements in the sound quality of hearing instruments are desirable to increase the use, benefit, acceptance and comfort of these instruments. Models of auditory perception can be used to assess sound quality by replacing time-consuming listening tests. In addition, developers can use model features to monitor the perceptual consequences of their algorithms. A computationally efficient model can also control hearing aid algorithms in real time by providing a running estimate on perception of signal-processing induced distortions.

While several instrumental measures of monaural audio quality are available, binaural and combined monaural and binaural measures are less well established. For the purpose of real-time algorithm control, computational efficiency is an additional requirement for such models. A contribution to this is presented in **chapter 5**. The physiologically plausible and perceptually validated binaural model, which was also the basis for the chapters 3 and 4, is extended to assess combined monaural and binaural audio quality. While its prediction accuracy can compete with previous, more complex models, the proposed model is mathematically simple and efficient. This makes it suitable for time-critical applications.

References

Abrams, HB., and Kihm, J. (**2015**). "An Introduction to MarkeTrak IX: A New Baseline for the Hearing Aid Market" .

Bennett, R. J., Laplante-Lévesque, A., Meyer, C. J., and Eikelboom, R. H. (**2018**). "Exploring Hearing Aid Problems: Perspectives of Hearing Aid Owners and Clinicians," Ear and Hearing **39**(1), 172, doi: `10.1097/AUD.0000000000000477`.

Brodie, A., Smith, B., and Ray, J. (**2018**). "The impact of rehabilitation on quality of life after hearing loss: A systematic review," European Archives of Oto-Rhino-Laryngology **275**(10), 2435–2440, doi: `10.1007/s00405-018-5100-7`.

Bronkhorst, and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," The Journal of the Acoustical Society of America **83**(4), 1508–1516, doi: `10.1121/1.395906`.

Brown, C. H. (**1994**). "Sound Localization," in *Comparative Hearing: Mammals*, edited by R. R. Fay and A. N. Popper, Springer Handbook of Auditory Research (Springer, New York, NY), pp. 57–96, doi: `10.1007/978-1-4612-2700-7_3`.

Committee on Accessible and Affordable Hearing Health Care for Adults, Board on Health Sciences Policy, Health and Medicine Division, and National Academies of Sciences, Engineering, and Medicine (**2016**). The National Academies Collection: Reports Funded by National Institutes of Health *Hearing Health Care for Adults: Priorities for Improving Access and Affordability* (National Academies Press (US), Washington (DC)).

Culling, J. F., and Lavandier, M. (**2021**). "Binaural Unmasking and Spatial Release from Masking," in *Binaural Hearing*, edited by R. Y. Litovsky, M. J. Goupell, R. R. Fay, and A. N. Popper, **73** (Springer International Publishing, Cham), pp. 209–241, doi: `10.1007/978-3-030-57100-9_8`.

Davis, A. C., and Hoffman, H. J. (**2019**). "Hearing loss: Rising prevalence and impact," Bulletin of the World Health Organization **97**(10), 646–646A, doi: `10.2471/BLT.19.224683`.

Derleth, P., Georganti, E., Latzel, M., Courtois, G., Hofbauer, M., Raether, J., and Kuehnel, V. (**2021**). “Binaural Signal Processing in Hearing Aids,” Seminars in Hearing **42**(3), 206–223, doi: `10.1055/s-0041-1735176`.

Dieudonné, B., and Francart, T. (**2019**). “Redundant Information Is Sometimes More Beneficial Than Spatial Information to Understand Speech in Noise:,” Ear and Hearing **40**(3), 545–554, doi: `10.1097/AUD.0000000000000660`.

Ferguson, M. A., Kitterick, P. T., Chong, L. Y., Edmondson-Jones, M., Barker, F., and Hoare, D. J. (**2017**). “Hearing aids for mild to moderate hearing loss in adults,” Cochrane Database of Systematic Reviews (9), doi: `10.1002/14651858.CD012023.pub2`.

Griffiths, T. D., Lad, M., Kumar, S., Holmes, E., McMurray, B., Maguire, E. A., Billig, A. J., and Sedley, W. (**2020**). “How Can Hearing Loss Cause Dementia?,” Neuron **108**(3), 401–412, doi: `10.1016/j.neuron.2020.08.003`.

Grothe, B., and Pecka, M. (**2014**). “The natural history of sound localization in mammals–a story of neuronal inhibition,” Frontiers in Neural Circuits **8**, 116, doi: `10.3389/fncir.2014.00116`.

Heffner, R. S. (**1997**). “Comparative Study of Sound Localization and its Anatomical Correlates in Mammals,” Acta Oto-Laryngologica **117**(sup532), 46–53, doi: `10.3109/00016489709126144`.

Hirsh, I. J. (**1948**). “The Influence of Interaural Phase on Interaural Summation and Inhibition,” The Journal of the Acoustical Society of America **20**(4), 536–544, doi: `10.1121/1.1906407`.

Jiang, F., Mishra, S. R., Shrestha, N., Ozaki, A., Virani, S. S., Bright, T., Kuper, H., Zhou, C., and Zhu, D. (**2023**). “Association between hearing aid use and all-cause and cause-specific dementia: An analysis of the UK Biobank cohort,” The Lancet Public Health **8**(5), e329–e338, doi: `10.1016/S2468-2667(23)00048-8`.

Lotfi, Y., Mehrkian, S., Moossavi, A., and Faghih-Zadeh, S. (**2009**). “Quality of life improvement in hearing-impaired elderly people after wearing a hearing aid,” Archives of Iranian Medicine **12**(4), 365–370.

Marquardt, D., Hohmann, V., and Doclo, S. (**2015**). “Interaural Coherence Preservation in Multi-Channel Wiener Filtering-Based Noise Reduction for Binaural Hearing Aids,” IEEE/ACM Transactions on Audio, Speech, and Language Processing **23**(12), 2162–2176, doi: `10.1109/TASLP.2015.2471096`.

Masterton, B., Heffner, H., and Ravizza, R. (**1969**). “The Evolution of Human Hearing,” The Journal of the Acoustical Society of America **45**(4), 966–985, doi: `10.1121/1.1911574`.

McCormack, A., and Fortnum, H. (**2013**). “Why do people fitted with hearing aids not wear them?,” International Journal of Audiology **52**(5), 360–368, doi: `10.3109/14992027.2013.769066`.

Olusanya, B. O., Neumann, K. J., and Saunders, J. E. (**2014**). “The global burden of disabling hearing impairment: A call to action,” Bulletin of the World Health Organization **92**, 367–373, doi: `10.2471/BLT.13.128728`.

Rennies, J., and Kidd, G. (**2018**). "Benefit of binaural listening as revealed by speech intelligibility and listening effort," The Journal of the Acoustical Society of America **144**(4), 2147–2159, doi: `10.1121/1.5057114`.

Seol, H. Y., and Moon, I. J. (**2022**). "Hearables as a Gateway to Hearing Health Care," Clinical and Experimental Otorhinolaryngology **15**(2), 127–134, doi: `10.21053/ceo.2021.01662`.

Shinn-Cunningham, B., Best, V., and Lee, A. K. C. (**2017**). "Auditory Object Formation and Selection," in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay, Springer Handbook of Auditory Research (Springer International Publishing, Cham), pp. 7–40, doi: `10.1007/978-3-319-51662-2_2;`.

Shinn-Cunningham, B. G. (**2005**). "Influences of spatial cues on grouping and understanding sound," in *ForumAcusticum*.

United Nations (**2015**). "Transforming our world: The 2030 agenda for sustainable development. In: Seventieth United Nations General Assembly, New York, 25 September 2015" .

Vaisberg, J. M., Beaulac, S., Glista, D., Macpherson, E. A., and Scollie, S. D. (**2021**). "Perceived Sound Quality Dimensions Influencing Frequency-Gain Shaping Preferences for Hearing Aid-Amplified Speech and Music," Trends in Hearing **25**, 2331216521989900, doi: `10.1177/2331216521989900`.

Wilson, B. S., and Tucci, D. L. (**2021**). "Addressing the global burden of hearing loss," The Lancet **397**(10278), 945–947, doi: `10.1016/S0140-6736(21)00522-5`.

World Health Organization (**2017**). "WHA70.13. World Health Assembly resolution on prevention of deafness and hearing loss: In: Seventieth World Health Assembly, Geneva, 31 May 2017. Resolutions and decisions, annexes. Geneva: ; 2017." .

World Health Organization (**2020**). "The top 10 causes of death" https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

World Health Organization (**2021**). "World report on hearing. Geneva" https://www.who.int/publications-detail-redirect/9789240020481.

World Health Organization (**2023**). "Deafness and hearing loss" https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

You, E., Lin, V., Mijovic, T., Eskander, A., and Crowson, M. G. (**2020**). "Artificial Intelligence Applications in Otology: A State of the Art Review," Otolaryngology–Head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery **163**(6), 1123–1133, doi: `10.1177/0194599820931804`.

CHAPTER 2

The model concepts

In the following, section 2.1 describes the foundation underlying this work. Section 2.2 introduces the interference concept applied in the chapters 3 and 4.

2.1 Concepts underlying this thesis

2.1.1 Binaural unmasking enabled by fluctuations in interaural phase

A tone presented with opposite phase to one ear compared to the other ear (i.e. antiphasic) can be detected in diotic noise (i.e. identically presented to both ears) at an about 15 dB lower signal-to-noise ratio (SNR) than a diotic tone. This effect is known as binaural unmasking (BU), the resulting shift in detection threshold is called binaural masking level difference (BMLD). A tone that has a constant interaural phase difference (IPD) of π added to a diotic noise causes fluctuations in the IPD of the mixed signal. This arises from he mixture of the amplitude fluctuations of the noise and the IPD of the tone (see Fig. 2.1). The strength of interaural fluctuations reflects the dissimilarity of left and right signals and that dissimilarity is reflected in a low coherence. While a diotic signal is perceived as a compact auditory event between the ears when listened to via headphones, an IPD

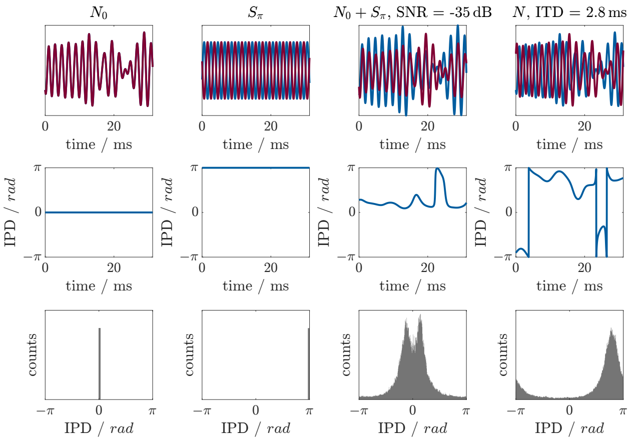

Figure 2.1: Basic interaural conditions and their IPD. Columns show a diotic noise, gammatone-filtered with a center frequency of 500 Hz, the same noise with an added antiphasic tone at SNR = -35 dB, and the same noise with an ITD of 0.4 ms. The rows show the waveforms, their instantaneous IPD, and the IPD histograms. It is visible that introducing an antiphasic tone or an ITD to an otherwise diotic noise causes IPD fluctuations.

causes a frequency-dependent lateralization, i.e. a shift towards one of the ears. Therefore, if the IPD and thus the laterality of a signal is changing rapidly over time, the perceptual consequence is a widening of the perceived auditory event, i.e. a less compact within-the-head representation. The laterality cue and the widening cue in combination enable binaural unmasking.

2.1.2 The delay-line model

The delay-line model as first postulated by Jeffress (1948) is one of the longest-standing models in sensory neuroscience (Encke and Dietz, 2022). For decades, it has been the basis for effectively describing binaural processing and especially for successfully explaining binaural unmasking (Durlach, 1963; Colburn, 1973, 1977). It assumes an array of neurons in the brainstem tuned to external delays in the range of a few milliseconds, i.e. ITDs, realized through axonal internal delays.

Besides a direct encoding of the stimulus ITD, this implies that a change in interaural correlation, caused by the addition of a tone to a noise, is equally well detectable for the whole range of noise ITDs. This can be seen as an internal compensation of the ITD, based on the idea that there are neurons having their maximum sensitivity at that specific ITD. This ITD-dependence of detectability has been characterized involving a function $p(\tau)$ reflecting the reduced compensation potency at higher internal delays τ (Colburn, 1973; Stern and Colburn, 1978). This is visualized in Fig. 2.3, left panel.

2.1.3 Filter bandwidth dictates binaural unmasking

With the bandpass filter bandwidth of ERB $=$ 79 Hz at 500 Hz center frequency (Patterson, 1976; Glasberg and Moore, 1990) which has been confirmed to well explain BU in delayed noise (Dietz *et al.*, 2021), the corresponding noise coherence already dictates the maximum achievable binaural unmasking (Langford and Jeffress, 1964; Rabiner *et al.*, 1966; Dietz *et al.*, 2021). This is because the noise coherence is determined by the spectrum, which is a general property of waves. Specifically, it is proportional to the inverse Fourier transform of the noise power spectral density, known as WIENER-KHINCHIN theorem (Wiener, 1930; Khintchine, 1934). Therefore, tone detection thresholds as both a function of noise ITD [left panel of Fig. 2.2, Langford and Jeffress (1964)] and of noise correlation [right panel of Fig. 2.2, Robinson and Jeffress (1963)] correspond well to the coherence of the noise after gammatone filtering with ERB $=$ 79 Hz. The consequence of the stimulus coherence resulting from basilar membrane filtering already determining detection thresholds is that there is no need and no room for a delay compensation plus its potency reduction $p(\tau)$. This is quantitatively modeled and discussed in Chapter 3. Therefore, all models presented in this thesis involve conventional peripheral filter bandwidths corresponding to Glasberg and Moore (1990).

2.1.4 The two-channel code

While axonal delays with a length as assumed by the delay-line concept have been found in barn owls (Carr and Konishi, 1988), such has not been found in mammals. Instead, it has been shown that the maximum firing rates of mammalian neurons in the medial superior olive (MSO) are limited to half the corresponding period, i.e. π (Marquardt and Mcalpine, 2007), and clustered around $\pi/4$ (McAlpine *et al.*, 2001). Therefore, from a physiologic standpoint, it is more reasonable to assume that binaural encoding is based on the relationship in activity of binaural nerons

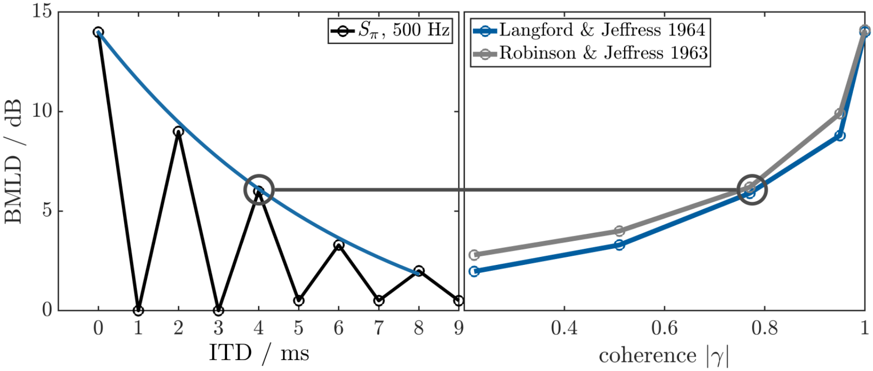

Figure 2.2: Illustration of the correspondence between noise coherence and BMLD; **left panel**: BMLD for S_π detection as a function of masker ITD (black circles connected by black line), redrawn from Langford and Jeffress (1964); **right panel**: BMLD for S_π detection as a function of noise interaural correlation, redrawn from Robinson and Jeffress (1963) (gray line). The horizontal dard gray line with circles illustrates the correspondence between the BMLD for a noise ITD of 4 ms and for a noise interaural correlation of 0.75.

between left and right hemispheres. This supports a rate-code instead of a delay line model (Encke and Dietz, 2022). However, in the past, two-channel models have not achieved the predictive power of delay-line models (Encke and Hemmert, 2018; Bouse *et al.*, 2019).

2.1.5 Mathematically efficient approximation: The complex correlation coefficient γ

Two orthogonal dimensions for effective binaural modeling

Based on previous models that evaluate IPD statistics (Goupell and Hartmann, 2006; Dietz *et al.*, 2008, 2021), Encke and Dietz (2022) presented a mathematically efficient, simplified formulation of the two-channel code. It interprets the best IPDs clustered around $\pm\frac{\pi}{4}$ as two correlation coefficients with $\frac{\pi}{2}$ phase offset. An important novelty of this approach is the orthogonality assumption, whereas the delay-line concept relies on the dependence of correlation units, reflected in weighting functions like centrality and straightness (Trahiotis and Stern, 1989). The orthogonality allows expressing the two coefficients as a complex number, termed the complex correlation coefficient γ, with the real and imaginary part representing the correlation coefficients in left and right hemispheres, respectively. Figure 2.3 illustrates the correspondence between the correlation function and the two orthogonal correlation coefficients. The complex formulation allows interpretation as a vector with the magnitude $|\gamma|$ representing coherence and the argument $\arg\{\gamma\}$ representing the phase angle between the two signals, i.e., the mean IPD (see Fig. 2.4). The γ model, first published by Encke and Dietz (2022), achieved a good predictive power for binaural detection. It directly interprets the Fisher-z (i.e. atanh) transformed difference of γ between the reference and the test signals as the ability of the models to discriminate between the two signals.

The delay-line and γ are equivalent in evaluating the interaural correlation coefficient ρ, i.e. the correlation at zero internal delay. However, the delay-line model adds an array of further correlation coefficients at a range of internal delays, adjusted by the mentioned $p(\tau)$ function, while the γ model adds a second, orthogonal dimension, i.e. the correlation coefficient associated with the second hemisphere.

The models presented throughout this theses build on the γ model because of the physiologic association and plausibility combined with mathematical efficiency and clarity.

Two orthogonal dimensions in other research fields

Another motivation to uptake γ is that adding a second or more orthogonal dimensions has historically led to more consistent and powerful models in various other fields of research. These include descriptions of waves and other phenomena. Some examples are:

Optics Interference and diffraction of waves are described in two orthogonal dimensions, expressed as complex numbers, as well as the polarization of light.

Social Sciences Nowaday's political landscapes have been argued to be more appropriately described involving two orthogonal dimensions, namely a liberitarian-authoritarian position and the conventional left-right economic position (Jackson and Jolly, 2021; Wagner *et al.*, 2023; Jolly *et al.*, 2022; Hooghe *et al.*, 2002). Representing these two orthogonal dimensions as a complex number allows encoding of, e.g., the political direction of a society: When averaging a number of unit vectors, reflecting individual positions, the angle of the resulting vec-

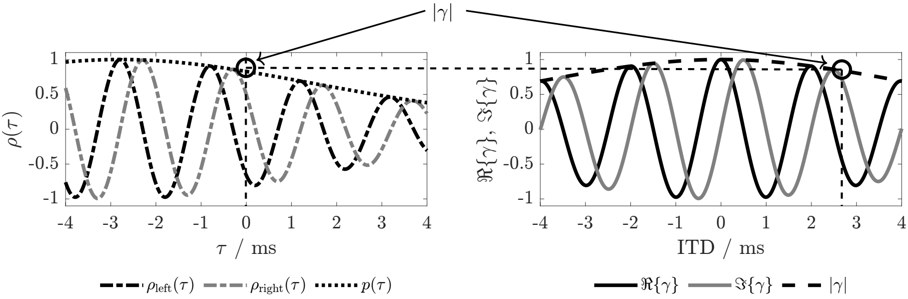

Figure 2.3: Illustration of the relationship of interaural correlation ρ and coherence γ. **Left panel:** Correlation functions for two orthogonal correlator units $\rho_{\text{left}}(\tau)$ and $\rho_{\text{right}}(\tau)$ associated with left and right hemispheres, as a function of the internal delay (or lag) τ, again for a noise with ITD = 2.8 ms, gammatone-filtered at 500 Hz with ERB = 79 Hz. The complex correlation coefficient γ is given by interpreting $\rho_{\text{left}}(\tau)$ and $\rho_{\text{right}}(\tau)$ as real part and imaginary part of a complex correlation coefficient ($\rho(\tau = 0)$). The real part $\Re\{\gamma\}$ is equal to the correlation coefficient $\rho(\tau = 0)$ and can be interpreted as the correlation coefficient encoded in one of the hemispheres, while the imaginary part $\Im\{\gamma\}$ then represents the correlation coefficient encoded in the other hemisphere, phase shifted by 90° = $\frac{\pi}{2}$. The magnitude $|\gamma|$ – the envelope – represents the coherence or phase predictability of the two underlying signals for a given ITD. A traditional delay-line based model would assume to evaluate the whole visible part of the correlation function and would adjust a function $p(\tau)$ to mimick the potency of the assumed delay compensation and, in case of a wider assumed filter, the temporal coherence decay. **Right panel:** γ plotted as a function of the noise ITD. The coherence for the noise ITD of 2.8 ms is now represented by the magnitude of γ at that ITD. The vector representation of γ is shown in Fig. 2.4.

tor reflects the resulting societal political direction. The magnitude can be interpreted as the coherence of the individual political positions.

2.2 Interference concept suggested in this thesis

As outlined in chapter 1, in both the spectral and the temporal domain, apparently larger analysis windows have been assumed in binaural compared to monaural hearing. However, in both domains there is also evidence for analysis windows being similarly small in binaural and monaural hearing. The starting point for this thesis is the following hypothesis: While a high-resolution system, i.e. smaller involved analysis windows, can potentially be unable to access its full resolution under certain conditions, the reverse is not possible. Therefore, the proposed concept is that the

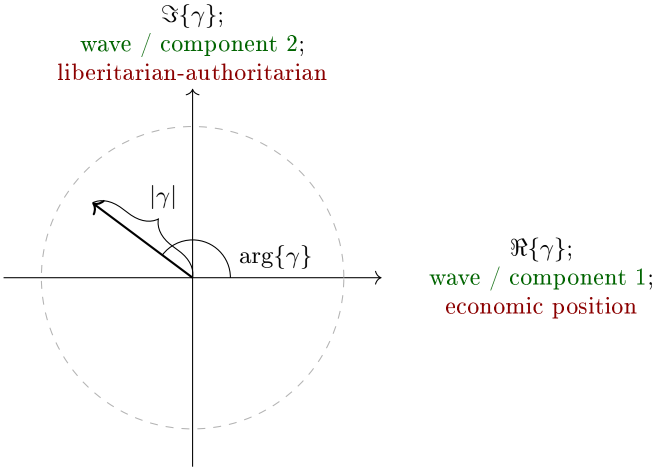

Figure 2.4: Exemplary complex space to illustrate how two-dimensional models are represented in the complex plane. The vector represents γ corresponding to the condition shown in Figs. 2.1 and 2.3. Examples from other research fields are: In optics, two waves or two light components are described using two orthogonal dimensions (green axis labels). As another example, two orthogonal dimensions have also been suggested to reflect the political landscape (red axis labels). The resulting vector represents the resulting two-dimensional description.

same, small analysis windows operate in binaural and monaural hearing. However, off-signal regions can interfere with signal detection and thus lead to the impression of an overall larger analysis window. The concept that interference can explain the apparently larger binaural analysis windows is inspired by crosstalk in electronics and adjacent-channel interference in radio systems. It is hypothesized that spectral or temporal off-signal changes in the interaural representation of a sound cannot always be ignored and thus affect the hearing sensation. This is based on previous observations in audition that have been associated with interference: For monaural hearing, detection of amplitude modulations have been described to depend on amplitude modulations at other frequencies. This has been termed modulation detection interference (Yost and Sheft, 1989; Bacon and Konrad, 1993; Mendoza *et al.*, 1995; Bernstein and Trahiotis, 1995; Oxenham and Dau, 2001). For binaural hearing, detection of changes in ITD or ILD have been shown to be impaired by a spectrally remote interferer with differing ITD/ILD, referred to as binaural inter-

ference. Both effects have been associated with both simultaneous and sequential perceptual grouping of the two sensations to one or two distinct streams (Oxenham and Dau, 2001; Best *et al.*, 2007). As perceptual grouping and streaming does not necessarily involve attention (Sussman *et al.*, 2007), this thesis models interference as a bottom-up process without involving top-down processing.
Specifically, the two cases where interference accounts for the apparent contradictions in binaural analysis windows are:

Spectral domain (chapter 3, Eurich *et al.* 2022) Detecting a pure tone in broadband masking noise varies with the frequency dependence of interaural coherence. The impaired detection for modulated masker coherence patterns is consistent with the assumption of larger binaural than monaural filter bandwidth. However, the non-impaired detection for frequency-independent masker coherence suggests that the same filter bandwidth determines binaural and monaural detection. Applying the proposed concept means to assume the generally accepted basilar membrane filter bandwidth plus a detrimental interference of lower coherence from off-signal bands. This work also resulted in a wave-form processing model which is published in the auditory modeling toolbox (AMT) version 1.4 (Majdak *et al.*, 2022).

Temporal domain (chapter 4, Eurich and Dietz 2023) Depending on interaural statistics in temporal surrounding of a target, its detection can be impaired. Analoguous to the spectral domain, this has been interpreted as longer temporal analysis window (“binaural sluggishness”). Other results can only be explained with the temporal resolution limited only by basilar membrane filtering, as is generally accepted for monaural hearing. Applying the proposed concept again means to assume a temporal resolution as resulting from basilar membrane filtering plus detrimental interference of interaural statistics across time, i.e. sluggish re-organization in case of rapid changes in interaural statistics.

References

Bacon, S. P., and Konrad, D. L. (**1993**). “Modulation detection interference under conditions favoring within- or across-channel processing,” The Journal of the Acoustical Society of America **93**(2), 1012–1022, doi: `10.1121/1.405549`.

Bernstein, L. R., and Trahiotis, C. (**1995**). “Binaural interference effects measured with masking-level difference and with ITD- and IID-discrimination paradigms,” The Journal of the Acoustical Society of America **98**(1), 155–163, doi: `10.1121/1.414467`.

Best, V., Gallun, F. J., Carlile, S., and Shinn-Cunningham, B. G. (**2007**). “Binaural interference and auditory grouping,” The Journal of the Acoustical Society of America **121**(2), 1070–1076, doi: `10.1121/1.2407738`.

Bouse, J., Vencovský, V., Rund, F., and Marsalek, P. (**2019**). “Functional rate-code models of the auditory brainstem for predicting lateralization and discrimination data of human binaural perception,” The Journal of the Acoustical Society of America **145**(1), 1–15, doi: `10.1121/1.5084264`.

Carr, C. E., and Konishi, M. (**1988**). “Axonal delay lines for time measurement in the owl’s brainstem,” Proceedings of the National Academy of Sciences **85**(21), 8311–8315, doi: `10.1073/pnas.85.21.8311`.

Colburn, H. S. (**1973**). “Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination,” The Journal of the Acoustical Society of America **54**(6), 1458–1470, doi: `10.1121/1.1914445`.

Colburn, H. S. (**1977**). “Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise,” The Journal of the Acoustical Society of America **61**(2), 525–533, doi: `10.1121/1.381294`.

Dietz, M., Encke, J., Bracklo, K. I., and Ewert, S. D. (**2021**). “Tone detection thresholds in interaurally delayed noise of different bandwidths,” Acta Acustica **5**, 60, doi: `10.1051/aacus/2021054`.

Dietz, M., Ewert, S. D., Hohmann, V., and Kollmeier, B. (**2008**). “Coding of temporally fluctuating interaural timing disparities in a binaural processing model based on phase differences,” Brain Research **1220**, 234–245, doi: `10.1016/j.brainres.2007.09.026`.

Durlach, N. I. (**1963**). “Equalization and Cancellation Theory of Binaural Masking-Level Differences,” The Journal of the Acoustical Society of America **35**(8), 1206–1218, doi: `10.1121/1.1918675`.

Encke, J., and Dietz, M. (**2022**). “A hemispheric two-channel code accounts for binaural unmasking in humans,” Communications Biology **5**(1), 1122, doi: `10.1038/s42003-022-04098-x`.

Encke, J., and Hemmert, W. (**2018**). “Extraction of Inter-Aural Time Differences Using a Spiking Neuron Network Model of the Medial Superior Olive,” Frontiers in Neuroscience **12**, doi: `10.3389/fnins.2018.00140`.

Eurich, B., and Dietz, M. (**2023**). “Fast binaural processing but sluggish masker representation reconfiguration,” The Journal of the Acoustical Society of America **154**(3), 1862–1870, doi: `10.1121/10.0021072`.

Eurich, B., Encke, J., Ewert, S. D., and Dietz, M. (**2022**). “Lower interaural coherence in off-signal bands impairs binaural detection,” The Journal of the Acoustical Society of America **151**(6), 3927–3936, doi: `10.1121/10.0011673`.

Glasberg, B. R., and Moore, B. C. (**1990**). “Derivation of auditory filter shapes from notched-noise data,” Hearing Research **47**(1-2), 103–138, doi: `10.1016/0378-5955(90)90170-T`.

Goupell, M. J., and Hartmann, W. M. (**2006**). “Interaural fluctuations and the detection of interaural incoherence: Bandwidth effects,” The Journal of the Acoustical Society of America **119**(6), 3971–3986, doi: `10.1121/1.2200147`.

Hooghe, L., Marks, G., and Wilson, C. J. (**2002**). “Does Left/Right Structure Party Positions on European Integration?,” Comparative Political Studies **35**(8), 965–989, doi: `10.1177/001041402236310`.

Jackson, D., and Jolly, S. (**2021**). “A new divide? Assessing the transnational-nationalist dimension among political parties and the public across the EU,” European Union Politics **22**(2), 316–339, doi: `10.1177/1465116520988915`.

Jeffress, L. A. (**1948**). “A place theory of sound localization.,” Journal of Comparative and Physiological Psychology **41**(1), 35–39, doi: `10.1037/h0061495`.

Jolly, S., Bakker, R., Hooghe, L., Marks, G., Polk, J., Rovny, J., Steenbergen, M., and Vachudova, M. A. (**2022**). “Chapel Hill Expert Survey trend file, 1999–2019,” Electoral Studies **75**, 102420, doi: `10.1016/j.electstud.2021.102420`.

Khintchine, A. (**1934**). “Korrelationstheorie der stationären stochastischen Prozesse,” Mathematische Annalen **109**(1), 604–615, doi: `10.1007/BF01449156`.

Langford, T. L., and Jeffress, L. A. (**1964**). “Effect of Noise Crosscorrelation on Binaural Signal Detection,” The Journal of the Acoustical Society of America **36**(8), 1455–1458, doi: `10.1121/1.1919224`.

Majdak, P., Hollomey, C., and Baumgartner, R. (**2022**). “AMT 1.x: A toolbox for reproducible research in auditory modeling,” Acta Acustica **6**, 19, doi: `10.1051/aacus/2022011`.

Marquardt, T., and Mcalpine, D. (**2007**). “A π-Limit for Coding ITDs: Implications for Binaural Models,” in *Hearing – From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer Berlin Heidelberg, Berlin, Heidelberg), pp. 407–416, doi: `10.1007/978-3-540-73009-5_44`.

McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). “A neural code for low-frequency sound localization in mammals,” Nature Neuroscience **4**(4), 396–401, doi: `10.1038/86049`.

Mendoza, L., Hall, III, J. W., and Grose, J. H. (**1995**). “Within- and across-channel processes in modulation detection interference,” The Journal of the Acoustical Society of America **97**(5), 3072–3079, doi: `10.1121/1.413105`.

Oxenham, A. J., and Dau, T. (**2001**). “Modulation detection interference: Effects of concurrent and sequential streaming,” The Journal of the Acoustical Society of America **110**(1), 402–408, doi: `10.1121/1.1373443`.

Patterson, R. D. (**1976**). “Auditory filter shapes derived with noise stimuli,” The Journal of the Acoustical Society of America **59**(3), 640–654, doi: `10.1121/1.380914`.

Rabiner, L. R., Laurence, C. L., and Durlach, N. I. (**1966**). “Further Results on Binaural Unmasking and the EC Model,” The Journal of the Acoustical Society of America **40**(1), 62–70, doi: `10.1121/1.1910065`.

Robinson, D. E., and Jeffress, L. A. (**1963**). “Effect of Varying the Interaural Noise Correlation on the Detectability of Tonal Signals,” The Journal of the Acoustical Society of America **35**(12), 1947–1952, doi: `10.1121/1.1918864`.

Stern, R. M., and Colburn, H. S. (**1978**). “Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position,” The Journal of the Acoustical Society of America **64**(1), 127–140, doi: `10.1121/1.381978`.

Sussman, E. S., Horváth, J., Winkler, I., and Orr, M. (**2007**). “The role of attention in the formation of auditory streams,” Perception & Psychophysics **69**(1), 136–152, doi: `10.3758/BF03194460`.

Trahiotis, C., and Stern, R. M. (**1989**). “Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase,” The Journal of the Acoustical Society of America **86**(4), 1285–1293, doi: `10.1121/1.398743`.

Wagner, S., Wurthmann, L. C., and Thomeczek, J. P. (**2023**). “Bridging Left and Right? How Sahra Wagenknecht Could Change the German Party Landscape,” Politische Vierteljahresschrift **64**(3), 621–636, doi: `10.1007/s11615-023-00481-3`.

Wiener, N. (**1930**). “Generalized harmonic analysis,” Acta Mathematica **55**(0), 117–258, doi: `10.1007/BF02546511`.

Yost, W. A., and Sheft, S. (**1989**). “Across-critical-band processing of amplitude-modulated tones,” The Journal of the Acoustical Society of America **85**(2), 848–857, doi: `10.1121/1.397556`.

CHAPTER 3

Lower interaural coherence in off-signal bands impairs binaural detection

This chapter is a formatted reprint of
Eurich, B., Encke, J., Ewert, S. D., and Dietz, M. (2022). Lower interaural coherence in off-signal bands impairs binaural detection. *The Journal of the Acoustical Society of America*, 151(6), 3927-3936, doi: `10.1121/10.0011673`.

Author contributions: BE, JE and MD developed the concept. BE implemented the model, computed the predictions, performed the analyses, prepared the figures and wrote as well as revised the manuscript. SE, JE and mainly MD participated in improving and revising the manuscript.

3.1 Abstract

Differences in interaural phase configuration between a target and a masker can lead to substantial binaural unmasking. This effect is decreased for masking noises with an interaural time difference (ITD). Adding a second noise with an opposing ITD in most cases further reduces binaural unmasking. Thus far, modeling of these detection thresholds required both a mechanism for internal ITD compensation and an increased filter bandwidth. An alternative explanation for the reduction is that unmasking is impaired by the lower interaural coherence in off-frequency regions caused by the second masker (Marquardt and McAlpine, 2009, JASA pp. EL177 - EL182). Based on this hypothesis, the current work proposes a quantitative multi-channel model using monaurally derived peripheral filter bandwidths and an across-channel incoherence interference mechanism. This mechanism differs from wider filters since it has no effect when the masker coherence is constant across frequency bands. Combined with a monaural energy discrimination pathway, the model predicts the differences between a single delayed noise and two opposingly delayed noises as well as four other data sets. It helps resolve the inconsistency that simulating some data requires wide filters while others require narrow filters.

3.2 Introduction

The detection of a pure tone in noise is facilitated by differences in the interaural phase between tone and noise (Hirsh, 1948). The improvement in the detection threshold compared to the diotic case is referred to as the binaural masking level difference (BMLD). The maximum BMLD is observed when detecting an antiphasic pure tone target (S_π) in an in-phase noise masker (N_0). Adding an interaural time difference (ITD) to the masker has been observed to reduce the BMLD (Langford and Jeffress, 1964). A particularly simple case is when the noise and the target tone have exactly opposite interaural phase differences. In this case, detection thresholds increase gradually and monotonically with increasing noise ITD (Rabiner *et al.*, 1966). The increase can be simulated accurately by exploiting changes in the cross-correlation coefficient of left and right signal after using a filter with an equivalent rectangular bandwidth (ERB) of 60 to 85 Hz at a center frequency of 500 Hz (Rabiner *et al.*, 1966; Dietz *et al.*, 2021). This bandwidth range resembles the established estimate of the human peripheral filter bandwidth obtained from monaural psychoacoustic experiments at this frequency which is 79 Hz (Glasberg and Moore, 1990) and referred to as standard filter bandwidth in the following.

Another explanation uses an array of different internal delays, known as delay lines (van der Heijden and Trahiotis, 1999; Stern and Colburn, 1978; Bernstein and Trahiotis, 2018, 2020). Jeffress (1948) suggested that the binaural system has the ability to compensate for the external ITD. The compensation accuracy or efficiency has been assumed to decrease with masker ITD in order to model the decreasing BMLD (Stern and Colburn, 1978; van der Heijden and Trahiotis, 1999; Bernstein and Trahiotis, 2017).

van der Heijden and Trahiotis (1999) generated a new stimulus which they termed "double-delayed noise" (diamonds in Fig. 3.1(E)) by adding two noises, one with a positive and one with a negative ITD. We refer to this as opposingly delayed noises (ODN). They found detection thresholds in ODN to be substantially higher than in "regular" delayed noise termed "single-delayed noise", SDN. Since internal delays can only compensate for the ITD of one noise, ODN limits the usefulness of the putative delay lines. Thus, van der Heijden and Trahiotis (1999) attributed the additional unmasking in SDN, compared to ODN, to the delay lines. Irrespective of the use of internal delays, however, a large part of the threshold differences between the two stimuli is caused by the interaural coherence oscillating as a function of ITD in ODN while monotonically decreasing in SDN (see the coherence ($|\gamma|$ pattern in Fig. 3.1(C)). More relevant for the role of internal delays are those ITDs that are the multiples of half the period. There, the coherence is the same in SDN and ODN but thresholds differ. For S_0 detection at 500 Hz this is the case at ITD = 1 ms and 3 ms, for S_π detection at ITD = 2 ms and 4 ms. So far, only the model of van der Heijden and Trahiotis (1999), which is based on delay lines, has precisely accounted for both SDN and ODN detection thresholds. The SDN-ODN detection threshold difference is therefore used as psychoacoustic evidence for several millisecond long delay lines (Stern *et al.*, 2019). The difference was in fact used to derive the length and potency of the delay line system (van der Heijden and Trahiotis, 1999).

However, two problems exist with establishing the psychoacoustically derived delay line length or internal delay distribution function. First, measured delays in binaural neurons of mammals are short compared to the respective period duration (McAlpine *et al.*, 2001; Joris *et al.*, 2006; see also Leibold and Grothe, 2015 for review) and thus too short to fulfil the lengths requirements of delay line models

(Thompson *et al.*, 2006; Marquardt and McAlpine, 2009; Stern *et al.*, 2019).

Second, if the delay-line models use their internal delays to account for SDN thresholds while correlation coefficient-based models (Rabiner *et al.*, 1966; Encke and Dietz, 2022) are equally precise for SDN without delay lines, the two model types must differ in some other manner, such as filter bandwidth. van der Heijden and Trahiotis (1999) used ODN thresholds to determine the filter bandwidth. They could best fit their ODN thresholds with filters of various shapes and an ERB of 130 to 180 Hz at 500 Hz center frequency. This is expectedly larger than what models without delay lines, such as Encke and Dietz (2022), required for SDN. The two versions cannot both be correct. Thus, either the SDN-threshold-based filter bandwidth is confounded by not considering delay lines or the ODN-based filter bandwidth fit by van der Heijden and Trahiotis (1999) is confounded by something else. For the latter, Marquardt and McAlpine (2009) offered a possible explanation. They identified the interaural coherence to be lower in certain off-frequency regions in ODN but not in SDN. They argued that the higher detection thresholds in ODN could also originate from some detrimental off-frequency impact related to the low coherence rather than from a wider filter bandwidth per se (upward arrow in Fig. 3.1(E)). If this is true, both SDN and ODN thresholds can potentially be predicted using the same standard filter bandwidth. Figure 3.1, panels (A), (B), and (D), show that the cross-power spectral density is constant across frequency in SDN but spectrally modulated in ODN (see Appendix for derivation).

Leaving aside the first physiologic argument, there are two options to account for the SDN-ODN difference, (1) wider filters combined with delay lines (downward arrow in Fig. 3.1(E)) or (2) filters with standard peripheral bandwidths and a detrimental off-frequency impact (upward arrow in Fig. 3.1(E)). However, recent data of SDN thresholds measured for different noise bandwidths can only be accurately simulated with filters falling into the standard peripheral bandwidth category (Bernstein and Trahiotis, 2020; Dietz *et al.*, 2021), causing a logical impasse for the wider-filters assumption even within the psychoacoustic domain and for SDN alone.

The aim of this study is thus to develop a model according to option (2) that accounts for SDN and ODN thresholds at the same time, using a standard filter bandwidth and – consequently – without several millisecond long delay lines. Rather, we suggest an across-frequency incoherence interference mechanism which is inspired by binaural interference (Bernstein and Trahiotis, 1995) and modulation detection interference (Yost and Sheft, 1989; Oxenham and Dau, 2001). With a low coherence qualitatively reflecting strong IPD fluctuations[1], this can be thought of as an interference of IPD fluctuations across frequency channels. With this mechanism, the same "hardware" causes different detection thresholds for maskers with the same on-frequency coherence but with a lower interaural coherence in off-frequency channels. The here developed incoherence interference will be described in Section 3.3 and used to predict critical binaural detection data in Section 3.4.

Besides the discussion concerning delay lines in humans and other mammals, the width of filters has caused an unresolved contradiction in the binaural literature that filters need to be narrow to account for some and broad to account for other data (see Verhey and van de Par, 2020 for a review). Generally speaking, detection thresholds in spectrally simpler maskers can be simulated using a standard peripheral filter bandwidth (Breebaart *et al.*, 2001b), whereas more complex maskers appear to be processed by wider filters or alternative across-frequency processes (Kolarik and Culling, 2010). We therefore evaluated our model with data from five different studies in three groups:

1. van der Heijden and Trahiotis (1999) combined all key aspects required to revisit Marquardt and McAlpine's hypothesis: (a) The SDN thresholds are planned to be determined by the decay of $|\gamma|$ with a 79 Hz-wide Gammatone filter. (b) The ODN thresholds supposedly will, despite the same 79 Hz on-frequency filter, be elevated by the across-channel incoherence interference.

2. Marquardt and McAlpine (2009) not only presented the above-mentioned hypothesis but also experimental data with a novel type of stimuli. There, SDN and ODN maskers are spectrally surrounded by bands that each have a different, constant IPD. Certain flanking-band IPDs do while others do not cause interaural incoherence at the transitions. Their reported differences impose a challenge for single-channel models that use a constant filter bandwidth.

3. Sondhi and Guttman (1966), Holube *et al.* (1998) and Kolarik and Culling (2010) reported detection thresholds of an S_π tone centered in an in-phase

---

[1] In contrast to measures of IPD fluctuations, such as the variance of the instantaneous IPD (Dietz *et al.*, 2021), both the interaural coherence $|\gamma|$ and the interaural cross-correlation function inherently weight the instantaneous IPD with the amplitudes. This is instrumental to quantitatively account for the masking of different noises with different statistics, such as low-noise noise or multiplicative noise. We therefore expect the present model to also account for detection thresholds obtained with such maskers.

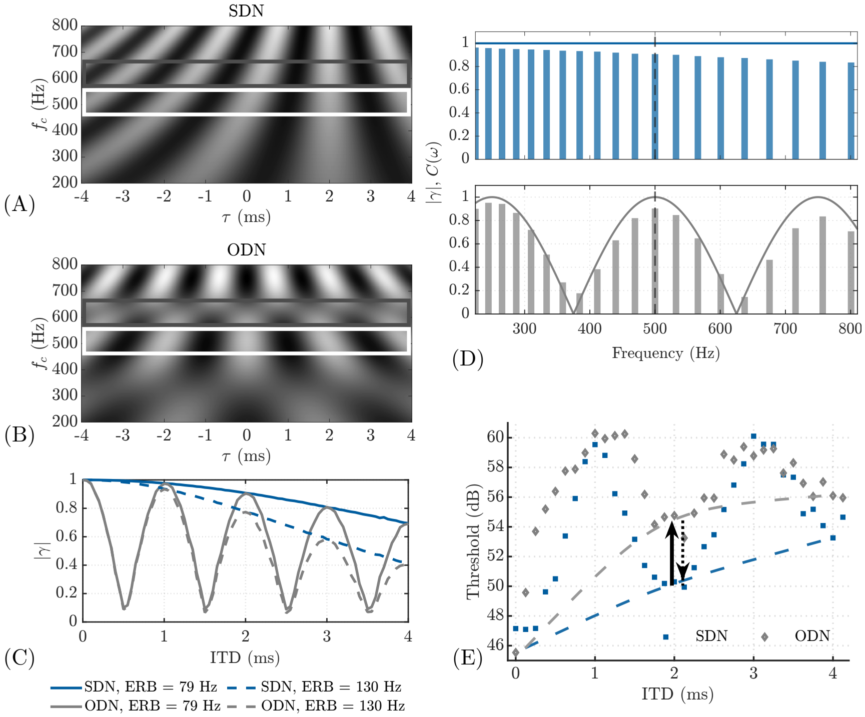

Figure 3.1: **(A)** Cross-correlogram of delayed noise (SDN) with ITD = 2 ms. White and black areas represent maxima and minima of the cross-correlation functions, respectively. The white box highlights the 500 Hz frequency channel while the gray box highlights a channel centered at 625 Hz. **(B)** Interaural cross-correlogram as in (A) but for opposingly delayed noises (ODN). **(C)** Interaural coherence $|\gamma|$ as a function of noise ITD for SDN (blue lines) and ODN (gray lines) for two underlying filter bandwidhts. **(D)** Continuous lines: Normalized cross-power spectral density (CPSD) at ITD = 2 ms as a function of frequency, $C(\omega)$, as derived in Eq. 3.11 et seq.; Bars: Interaural coherence $|\gamma|$ of the signals after peripheral Gammatone filtering. **(E)** Thresholds of S_π detection in SDN and ODN as a function of ITD from van der Heijden and Trahiotis (1999). The dashed lines symbolize the coherence-decline-induced threshold increase determined by a filter bandwidth of ERB = 79 Hz (lower line) and ERB = 130 Hz (upper line). As denoted by the arrows, the data can be explained in two ways: (1: dotted downward arrow) The ODN thresholds are determined by the cross-correlation function at 500 Hz and a bandwidth $\geq$ 130 Hz. A delay line causes the lower SDN thresholds. (2: solid upward arrow) The SDN thresholds are determined by the ITD-dependent coherence as derived from an ERB of 79 Hz. Off-frequency incoherence in ODN causes higher ODN thresholds.

noise that is spectrally surrounded by antiphasic noise. These simulations are included for an additional discussion about the proposed standard-filter-plus-off-frequency-impact concept, since larger binaural than peripheral bandwidths have previously been derived based on such data.

3.3 Description of the Model

Figure 3.2 shows the processing stages of the proposed model. It is designed as a numerical multi-channel model through all stages, but these were here realized and tailored to predict binaural-detection data with a 500 Hz pure-tone target. The model builds on the analytical single-channel model approach of Encke and Dietz (2022). It furthermore includes an across-frequency incoherence interference mechanism. It consists of a multi-channel binaural processing pathway and a monaural pathway in order to account both for conditions that provide interaural or only energetic cues. In both pathways, multiple tokens of the processed representation of the condition-specific masker only are compared to the representation of signal plus masker. This comparison has been suggested to mimic a subject's strategy of comparing a stimulus to a learned reference template (Breebaart *et al.*, 2001a; Bernstein and Trahiotis, 2017). Based on these comparisons, both pathways deliver a sensitivity index (d'). An optimal combination of the pathways' estimates gives the overall d' estimate of the model (Green, 1966; Biberger and Ewert, 2017).

3.3.1 Peripheral Processing

The left and right input signals were processed with a fourth-order Gammatone filterbank that represents basilar-membrane bandpass filtering. The filterbank implementation by Hohmann (2002) was employed with a spacing of five filters per ERB in the range of 67 Hz to 1000 Hz. The grid was defined by centering one filter at 500 Hz. This filter had an ERB of 79 Hz (Glasberg and Moore, 1990) and was indexed with $k = 0$.

To focus on the impact of the spectral masker properties discussed above, the present implementation did not include any other peripheral processing such as low-pass filtering, power-law compression or half-wave rectification. Only Gaussian noise was used as masker and only 500-Hz tones as targets.

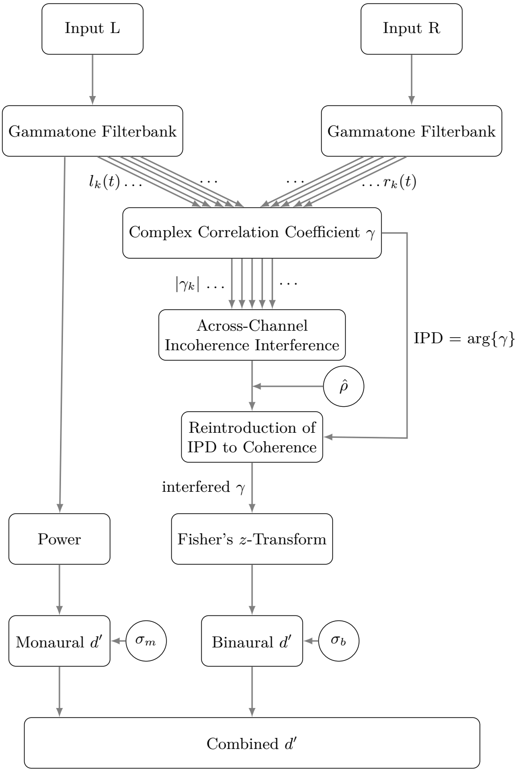

Figure 3.2: Processing stages of the proposed model. See main text for details.

3.3.2 Binaural Pathway

The correlation coefficient $\gamma(\tau = 0) = \gamma$ was derived from the analytical (i.e. complex-valued) left and right signals $l(t)$ and $r(t)$ in the frequency channel k, provided by the Gammatone filterbank:

$$\gamma_k = \frac{\overline{l_k(t)^* r_k(t)}}{\sqrt{\overline{|l_k(t)|^2}\,\overline{|r_k(t)|^2}}} \tag{3.1}$$

where $\overline{\bullet}$ denotes the mean over the duration of the signal. This results in one complex correlation coefficient per frequency channel, averaged over the whole stimulus duration. The complex-valued correlation coefficient was used because it conveniently combines information about both the mean IPD as $\arg\{\gamma\}$ and about the interaural coherence $|\gamma|$. While the Introduction mentioned a mismatch between mammalian physiology and delay line models, it should be noted that the seemingly abstract use of complex-valued correlation is identical to two real-valued correlations with a 90° phase offset. Such two orthogonal correlators exist in the form of the average left- and right hemispheric binaural neuron in mammals (McAlpine *et al.*, 2001; Joris *et al.*, 2006). The physiologic relation of γ is explained in more detail in Encke and Dietz (2022).

As pointed out in the Introduction, the novelty of the present model is the incoherence interference across frequency channels. The term *incoherence interference* was chosen to describe the purely detrimental nature of the interaction. Only channels with lower coherence affect their neighborhood, but not the other way around. This process is implemented as a *restricted* across-channel weighted average of the coherence $|\gamma|_k$: The $|\gamma|_k$ are limited such that they can no more exceed the on-frequency $|\gamma|$, thus referred to as $|\gamma|_{k,\mathrm{lim}}$.

$$|\gamma|_w = \sum_{-m}^{m} w(k)|\gamma|_{k,\mathrm{lim}} \tag{3.2}$$

$w(k)$ symbolizes a function that weights the contribution of a channel k to the resulting $|\gamma|_w$. The employed weighting function has an exponential decay described by

$$w(k) = e^{-|k|/(b\sigma_w)}. \tag{3.3}$$

σ_w represents the decay parameter, normalized by the number of filters per ERB, b. The double-exponential decay shape was chosen by empirical trials. While the exact shape of the window was not crucial, we did not obtain more precise simulations with other shapes.

For a masker coherence close to zero or at the practically irrelevant case of a positive signal-to-noise ratio (SNR), adding a target with an IPD of π relative to the masker can swap the mean IPD from the masker to that of the target. In special cases, the masker alone and masker plus target can have the same coherence but differ in their mean IPD and thus in their correlation. Hence, the interaural coherence $|\gamma|$ is not sufficient as a decision variable. Instead, γ, including both coherence and the mean IPD, is required. Therefore, the original mean IPD is now reintroduced to the coherence after the limitation and interference stage, so that the model can operate on the *complex* correlation coefficient as suggested by Encke and Dietz (2022).

$$\gamma_w = |\gamma|_w e^{\arg\{\gamma_0\}} \tag{3.4}$$

Unity-limited measures such as coherence or correlation can be Fisher z (i.e. atanh) transformed for the purpose of variance normalization (McNemar, 1969; Just and Bamler, 1994), as often applied in psychophysics (e.g., Lüddemann *et al.*, 2007; Bernstein and Trahiotis, 2017). As in Encke and Dietz (2022), γ_w is multiplied by a model parameter $\hat{\rho} < 1$ to avoid an infinite sensitivity to deviations from a coherence of unity. This is equivalent to adding uncorrelated noise to the two input signals. The decision variable of the binaural pathway is thus

$$\zeta = z[\hat{\rho}\gamma_w] \tag{3.5}$$

where $z[\bullet]$ is the Fisher z-transform applied to the modulus of γ_w while leaving the argument unchanged.

In the signal detection stage, the d' is obtained based on the difference between the ensemble averages of the representations of the target signal plus noise, ζ_{N+S}, and the representations of the noise alone, ζ_N:

$$d'_b = \frac{|\zeta_{N+S} - \zeta_N|}{\sigma_b} \tag{3.6}$$

The internal noise σ_b defines the sensitivity of the binaural model pathway (Dietz *et al.*, 2021).

3.3.3 Monaural Pathway

For the monaural pathway, the power P of the on-frequency filter channel was evaluated. It is half the squared mean of the envelope across the whole signal duration (Biberger and Ewert, 2016). The envelope is the modulus of the complex-valued filter output:

$$P = \frac{\overline{|u_0(t)|^2}}{2}. \tag{3.7}$$

In the stimuli employed in this study, the power is identical in the left and right channels, thus it is sufficient to evaluate only one side.
For a signal-induced power change $\Delta P = P_{N+S} - P_N$, the processing accuracy is limited by a level-dependent internal noise with a Gaussian distribution of amplitudes and a standard deviation of σ_m. Thus, the sensitivity of the monaural pathway is equivalent to

$$d'_m = \frac{\Delta P / P_{\text{avg}}}{\sigma_m}, \tag{3.8}$$

where P_{avg} represents the average power between P_{N+S} and P_N.

3.3.4 Detector

The sensitivity indices of the binaural, d'_b, and monaural pathway, d'_m, were combined assuming two independent information channels (Green, 1966; Biberger and Ewert, 2016),

$$d'_{b+m} = \sqrt{d'^2_b + d'^2_m}. \tag{3.9}$$

The d' that corresponds to the experiment-specific detection thresholds was obtained via table-lookup (Numerical evaluation in Hacker and Ratcliff, 1979). This depends on the number of intervals as well as the specific staircase procedure used in the simulated experiments. For each condition, the model was evaluated for a range of target levels. This delivered the psychometric function. The predicted detection threshold was obtained from a straight line fitted to the logarithmic d'. The model parameters were manually adjusted in order to optimize the prediction accuracy. The resulting parameter values are given in Table 3.1.

3.4 Predictions of Binaural-Detection Datasets

In all experiments, a 500 Hz S_π or S_0 tone was to be detected in a broadband Gaussian noise masker. Figures 3.3, 3.4, and 3.5 show the experimental data denoted by symbols, the predictions of the proposed model including incoherence interference (continuous lines), as well as excluding incoherence interference (as dotted lines). Three types of binaural-detection experiments were simulated, as described in detail in the following subsections. Table 4.1 summarizes the parameter values used to simulate the experimental conditions. It further lists the non-adjusted coefficient of determination (R^2, interpretable as the proportion of variance in the data explained by the model) and the root-mean-square error (RMSE) of the simulations both with and without the proposed incoherence interference.

Experiment	Signal	Variable	$\hat{\rho}$	σ_b	σ_w	σ_m	with R^2	with RMSE / dB	without R^2	without RMSE / dB
van der Heijden and Trahiotis (1999)	π	ITD	0.91	0.20	0.50	0.40	0.94	0.85	0.78	1.45
	0		0.86	0.17	0.65	0.40	0.87	0.86	0.57	1.38
Marquardt and McAlpine (2009)	0	BW	0.89	0.24	0.65	0.40	0.96	0.37	-0.62	1.97
Kolarik and Culling (2010)	0	BW	0.91	0.20	0.50	0.40	0.97	0.67	0.42	3.07

Table 3.1: Summary of the simulated experiments and predictions. *Columns 1 - 3*: Simulated experiment, IPD of the used target signal, independent variable. *Columns 4 - 7*: Used model parameters: $\hat{\rho} < 1$: Maximum coherence (internal noise); σ_b: Standard deviation of the internal noise to determine the absolute performance of the binaural pathway; σ_w: Slope parameter of the double-exponential across-channel interaction window (normalized by the number of filters per ERB); σ_m: Standard deviation of the level-dependent internal noise to determine the accuity of the monaural pathway; *Columns 8 - 11*: Accuracy of the predictions with and without incoherence interference: Coefficient of determination (R^2); root-mean-square errors (RMSE) of the predictions.

3.4.1 van der Heijden & Trahiotis 1999

In this arguably most central experiment, detection thresholds of an S_0 target tone (Fig. 3.3, upper panel) as well as of an S_π tone (Fig. 3.3, lower panel) were measured as a function of the interaural masker ITD in steps of 0.125 ms. The bandwidth of the masker was 900 Hz. As outlined in the Introduction, the ODN consisted of two superimposed noises with opposite ITD. The experiment performed by van der Heijden and Trahiotis (1999) employed a four-interval, two-alternative forced choice task (4I-2AFC, first and fourth intervals always contained only the masker and served as queuing intervals). Their adaptive 2-down 1-up stair case procedure estimated the 70.7 % correct-response threshold. This is equivalent to a d' of 0.78 at threshold. Thus, as described in Section 3.3.4 the model determined the threshold in the form of the signal level producing this d'. The continuous lines in Fig. 3.3 show the simulations of the presented model, including the across-channel incoherence interference. From visual inspection, the simulations captured all effects from the experimental thresholds and the critical threshold differences between SDN and ODN at all ITDs under both conditions. Specifically, the critical threshold differences of 3.5 dB at an ITD = 1 ms in the S_0 condition and 4 dB at an ITD = 2 ms in the S_π condition are precisely accounted for. This good correspondence is also reflected in the

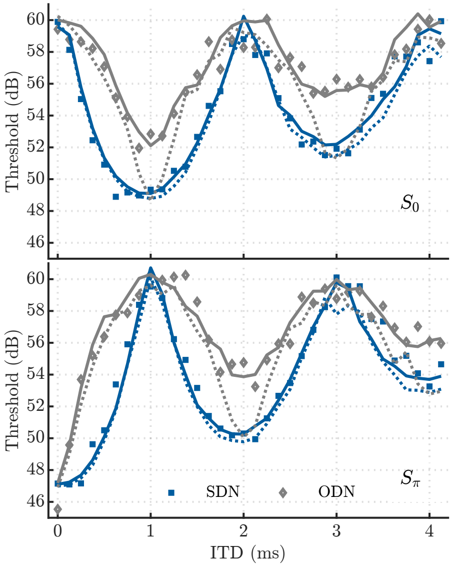

Figure 3.3: Experimental data from van der Heijden and Trahiotis (1999) (symbols). The continuous lines show the predictions of the presented model including the across-channel incoherence interference. The dashed lines show predictions for ODN excluding interference (single-channel version), equivalent to Encke and Dietz (2022). *Upper panel:* Detection thresholds with S_0 target; *lower panel:* S_π target.

around 90 % explained variance under both conditions and RMS errors of less than 1 dB. The dashed lines show simulations excluding the across-channel incoherence interference (single-channel model, cf. Encke and Dietz, 2022) but all other model parameters unchanged. This shows that a large amount of the threshold differences is already explained by differences in the on-frequency coherence. As mentioned in the Introduction: In much the same way as ODN coherence oscillates as a function of analysis frequency (Fig. 3.1(D)), it also fluctuates as a function of the masker ITD (Fig. 3.1(C)). Particularly at ITD = 0.5 ms, ODN is incoherent in the 500-Hz band, whereas SDN is almost fully coherent. This, and not the across-frequency process, causes the difference in the simulated thresholds at this ITD. The across-frequency process only comes into play at those ITDs where the coherence at 500 Hz (on-frequency) is nearly identical in SDN and ODN (upper panel: ITD = 1 ms and 3 ms; lower panel: ITD = 2 ms and 4 ms).

3.4.2 Marquardt & McAlpine 2009

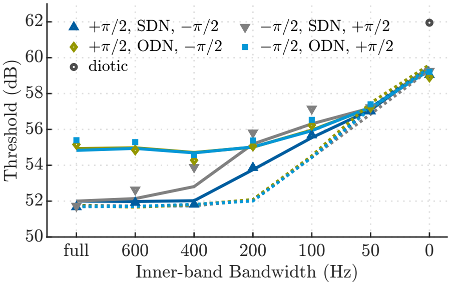

Figure 3.4: Experimental data from Marquardt and McAlpine (2009, symbols) and model predictions (lines). Detection thresholds are given as function of the inner-band bandwidth. The inner band contains delayed noise (triangles) or opposingly delayed noises (diamonds and bullets) with a fixed ITD = 1 ms while the flanking bands have a constant IPD of $+\pi/2$ (upward triangle and diamond) and $-\pi/2$, or vice versa (downward triangle and bullet). Continuous and dashed lines again show predictions with and without across-frequency incoherence interference, respectively.

The masker of this experiment contained SDN and ODN centered at the frequency of the S_0 target tone with a constant ITD in the inner band. The inner band was spectrally surrounded by bands that each had a constant IPD of $\pi/2$ and $-\pi/2$, or vice versa. Thresholds are given as a function of the inner-band bandwidth. The resulting phase transitions between inner and flanking bands have been hypothesized to impair the detection if they cause a frequency region of low interaural coherence. The lower and upper frequency limits of the composite stimuli are 50 Hz and 950 Hz, respectively. The two-interval-two-alternative-forced choice task with a 3-down 1-up procedure that was used estimated the thresholds to be 79.4 % correct. This corresponds to $d' = 1.14$ at the threshold predicted by the model. In Fig. 3.4, detection thresholds of the S_0 tone are shown as a function of the bandwidth of the inner band. Again, the model predicted all critical characteristics of the data. These include the 3 dB difference between SDN and ODN at the full inner-band bandwidth (same as ITD = 1 ms in the S_0 condition in van der Heijden and Trahiotis, 1999), the elevated SDN thresholds in the [-$\pi/2$, SDN, +$\pi/2$] compared to the [+$\pi/2$, SDN, -$\pi/2$] condition and the 3 dB BMLD where the inner-band bandwidth is zero. Without the incoherence interference, the predictions cannot be distinguished between the different conditions of the experiment. They deviate more from the mean than the data, resulting in a negative R^2.

3.4.3 Experiments on the operating bandwidth in binaural detection

Several studies investigated the operating bandwidth in binaural detection using maskers that contain two flanking bands which differ in their interaural configuration from the inner band (Sondhi and Guttman, 1966; Holube *et al.*, 1998; Kolarik and Culling, 2010). The masking noise is diotic (N_0) in the inner band and antiphasic (N_π) in the flanking bands. Detection thresholds of an S_π target tone were again measured as a function of the inner-band bandwidth. Results are expressed as the difference between thresholds in the flanked condition and the threshold without an inner band, i.e. $N_\pi S_\pi$. In Fig. 3.5, the circles mark the threshold differences reported by Kolarik and Culling (2010, centered condition), which represent averages across their three participants. The triangles show individual thresholds of the two participants in the study by Holube *et al.* (1998, rectangular condition). The gray diamonds show the data from Sondhi and Guttman (1966). Our model predictions were oriented on the 2-down 1-up 2I-2AFC paradigm employed in Kolarik and Culling (2010), equivalent to $d' = 0.78$ at threshold. The black continuous line shows the model predictions with the same parameter settings (see Table 3.1) as used to predict the S_π detection thresholds in van der Heijden and Trahiotis (1999) (see our predictions in Fig. 3.3(B)). The dotted black line shows model predictions without the across-channel incoherence interference, so that detection was purely determined by the ERB = 79 Hz Gammatone filter centered at 500 Hz. Despite the large deviations between and within experiments, the model predictions involving the incoherence interference captured the shape of the decreasing thresholds with increasing inner-band bandwidth.

3.5 Discussion

In this study, those binaural detection thresholds that previously have underpinned the psychoacoustic necessity of several millisecond long delay-lines were simulated involving across-frequency incoherence interference and monaurally derived peripheral filter bandwidths.

As long as the masker coherence is fairly constant across frequency bands, experi-

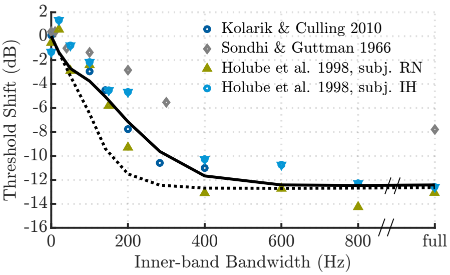

Figure 3.5: Symbols denote data from binaural detection experiments with the configuration $N_{\pi 0 \pi} S_{\pi}$ as a function of the inner-band (N_0) bandwidth; continuous and dotted line: Model prediction with and without across-incoherence incoherence interference, respectively.

ments on binaural detection can be explained purely on the basis of the coherence $|\gamma|$ defined by a 79 Hz wide Gammatone filter at $f_c = 500\,\text{Hz}$ (Rabiner *et al.*, 1966; Encke and Dietz, 2022). This includes fully coherent broadband noise maskers (Hirsh, 1948; van de Par and Kohlrausch, 1999), mixtures of correlated and uncorrelated noise (Robinson and Jeffress, 1963; Pollack and Trittipoe, 1959; Bernstein and Trahiotis, 2014), and experiments where the interaural coherence of the masker is reduced by an ITD (Langford and Jeffress, 1964; Rabiner *et al.*, 1966; Bernstein and Trahiotis, 2020). However, the on-frequency coherence does not account for thresholds obtained with maskers where these properties change substantially across filter bands. Specifically, the single-channel model version as proposed in Encke and Dietz (2022) is neither able to predict all of the threshold differences between SDN and ODN nor experiments like Marquardt and McAlpine (2009) and Kolarik and Culling (2010) that involve IPD transitions in the masker spectrum (see dashed lines in Figs. 3.3, 3.4, 3.5, as well as the corresponding R^2 and RMSE given in Table 3.1).

Marquardt and McAlpine (2009) hypothesized across-channel processing in the binaural system to explain the reduced binaural benefit under such conditions. Here, we extended the analytical model by Encke and Dietz (2022) to a multi-channel numerical signal-processing model with incoherence interference. The proposed model differs from approaches assuming wider binaural filters (e.g. van der Heijden and Trahiotis, 1999; Kolarik and Culling, 2010), as, for example, wider filters reduce the interaural coherence of SDN, whereas incoherence interference does not reduce it. For stimuli with spectrally constant coherence and masker-target phase relations, like SDN and all conditions simulated by Encke and Dietz (2022), the incoherence interference has no effect and the model operates on the standard filter bandwidths of its peripheral filterbank. Modeling an interference process, our approach also differs from the symbolic model suggested by Marquardt and McAlpine (2009), which sums interaural cues after binaural interaction. Their implementation is also different from wider filters but still causes a stronger damping of binaural sensitivity with increasing masker ITD, which is not seen in the data.

The proposed concept of a detrimental incoherence interference is comparable to modulation detection interference, as shown and discussed by, e.g., Yost and Sheft (1989) and Oxenham and Dau (2001). Similar to the proposed across-channel incoherence interference, this is modeled by modulation patterns interacting across channels, while energetic spectral masking properties are spectrally limited by peripheral filters (Piechowiak *et al.*, 2007; Dau *et al.*, 2013). Furthermore, a similar process is thought to underlie binaural interference as observed by, e.g., Bernstein and Trahiotis (1995); Best *et al.* (2007); McFadden and Pasanen (1976).

The dataset of van der Heijden and Trahiotis (1999) contains both SDN and ODN and is therefore the critical challenge for binaural detection models[2]. Both van der Heijden and Trahiotis' and our model simulate the data very accurately. Therefore, the discussion focuses on consequences and plausibility of the two different concepts.

The bandwidth of the signals immediately prior to binaural interaction dictates the temporal coherence and thus the decline of BMLD with increasing noise ITD in the

---

[2] The most comprehensive simulation of dichotic tone in noise detection thresholds using a cross-correlation-based model is by Bernstein and Trahiotis (2017). It is not expected to simulate the ODN detection thresholds of van der Heijden and Trahiotis (1999) with a good accuracy, because an ERB of at least 130 Hz is necessary. Other ODN stimuli, used experimentally by Bernstein and Trahiotis (2015), were included in the model test battery by Bernstein and Trahiotis (2017). Those ODN stimuli, however, differed in several ways from the former. First, the target frequency is 250 Hz, compared to 500 Hz in van der Heijden and Trahiotis (1999) and in all other studies here simulated. Second, instead of fixing the target tone to S_0 or S_π, the target is delayed by the same amount as one of the two noises, i.e. $(N_0)_{\pm\mathrm{ITD}}(S_\pi)_{\mathrm{ITD}}$. Such an approach is useful for SDN, as it ensures a constant π difference between the IPDs of the noise and of the tone. For ODN, however, the IPD of the second noise relative to the tone is offset from π by $2\times$ITD. This type of stimulus therefore causes an even more complex ITD-dependence of threshold, which offers no advantage over the ODN from van der Heijden and Trahiotis (1999) for filter estimation. With both definitions, corresponding SDN and ODN stimuli can be generated only if the ITD is an integer or a half-integer multiple of the target period (i.e. $\mathrm{ITD} = n/2f$, $n \in \mathbb{N}$). In Bernstein and Trahiotis (2015), Fig. 1, Panel a), these are the two data points at ITD = 2 and 4 ms. SDN and ODN thresholds are, however, very similar at those points. Third, the masker bandwidth is 50 Hz. For such a masker bandwidth smaller than the peripheral filter width, neither van der Heijden and Trahiotis (1999) nor our model would predict a considerable threshold difference between SDN and ODN at an ITD of 2 and 4 ms, since there are no off-frequency regions of considerably lower coherence.

absence of internal ITD compensation (Langford and Jeffress, 1964; Rabiner *et al.*, 1966; van der Heijden and Trahiotis, 1999; Dietz *et al.*, 2021). To date, two of the arguably most comprehensive datasets of dichotic tone-in-noise detection, van der Heijden and Trahiotis (1999) and Bernstein and Trahiotis (2020), have self-reported mutually exclusive requirements for the filter bandwidth (ERB = 130...180 Hz vs. ERB $\leq$ 100 Hz at 500 Hz).

A variety of studies aim to estimate the bandwidth at the binaural input stage by means of dichotic tone-in-noise detection, but no consistent picture emerges. There is, for example, a difference in estimated bandwidth between band-widening and notched-noise BMLD data, and between stimuli with different flanking bands (e.g., Kolarik and Culling, 2010). Particularly this stimulus-type dependence of the "apparent bandwidth" challenges the assumption that all stimuli are processed by the same filters. To us, the most reasonable "unifying" explanation is that filter properties arise from the basilar membrane and also the binaural system can make full use of this spectral resolution. The observation that there is less spectral resolution in some cases is then best explained by an across-frequency process for certain stimulus features – but in contrast to wider filters it is not affecting all features. The proposed incoherence interference may be this missing across-frequency process. At least it appears to reduce or even eliminate inconsistencies in estimating the bandwidth from various binaural detection experiments.

Another mechanism which has been proposed in the context of bandwidth estimation is an optimal combination of target detectability across frequency channels. Masking patterns in dichotic band-widening experiments have a knee-point at larger bandwidths than their diotic counterparts (van de Par and Kohlrausch, 1999; Bourbon and Jeffress, 1965). van de Par and Kohlrausch (1999) hypothesized that in narrowband maskers, the similar SNR across frequency channels can be exploited to reduce masking. A model which includes such a mechanism (Breebaart *et al.*, 2001b) accounts for the band-widening masking pattern using standard filter bandwidths (i.e. bandwidths as proposed by Glasberg and Moore, 1990). It also correctly predicts that the knee-point is only shifted to a higher bandwidth if the masker is fully or almost fully correlated (van der Heijden and Trahiotis, 1998).

Most recent binaural models, such as Bernstein and Trahiotis (2017) and Encke and Dietz (2022) already assume a bandwidth as narrow as the peripheral bandwidth. This is also in line with direct measurements of the bandwidth in ITD-sensitive inferior colliculus neurons in cats by Mc Laughlin *et al.* (2008). For delayed noise, as used by van der Heijden and Trahiotis (1999), they found that damping of the cross-correlation function corresponds to the peripheral bandwidth at the respective center frequency.

With the present implementation, the binaural pathway parameters ($\hat{\rho}$, σ_b, σ_w) had to be adjusted slightly between conditions with S_π targets and conditions with S_0 targets (see Table 3.1). This is due to the binaural system's sensitivity depending on the baseline IPD (Hirsh, 1948). An angular compression of the decision variable space $\{\zeta\}$ at large IPDs is a possible model extension. Delay-line models can account for this dependence with a corresponding $p(\tau)$ function which defines the sensitivity of the model as a function of its internal delay. However, they then incorrectly predict better unmasking with $N_{ITD}S_0$ compared to $N_\pi S_0$ when ITD $= T/2$ (Breebaart *et al.*, 1999). Simulating the data of Marquardt and McAlpine (2009) required slightly different parameter values because their listeners obtained different thresholds compared to van der Heijden and Trahiotis (1999) for identical stimuli. This may be due to the different number of presented intervals. Identical model parameters were used for the S_π conditions of van der Heijden and Trahiotis (1999) and Kolarik and Culling (2010).

3.6 Conclusion

Interaural incoherence interference enables the presented binaural model to simulate detection thresholds both for maskers with a spectrally constant and with a spectrally modulated coherence. Employing auditory filters with monaurally estimated bandwidth Glasberg and Moore (1990), it predicts the reduced unmasking in opposingly-delayed noises (van der Heijden and Trahiotis, 1999) compared to regular delayed noise. The concept can help to resolve the inconsistency that binaural models require filter bandwidths as estimated monaurally for most data sets (Bernstein and Trahiotis, 2017, 2020), but at least 1.6 times wider filters for broadband opposingly delayed noises van der Heijden and Trahiotis (1999) and other spectrally complex maskers Verhey and van de Par (2020). The main consequence of using a standard filter bandwidth is that the decline of the binaural benefit with masker ITD can be simulated without internal ITD compensation, as first suggested by Langford and Jeffress (1964).

3.7 Acknowledgments

We thank Steven van de Par for fruitful discussions and comments on the manuscript as well as Matthew J. Goupell and two anonynomous reviewers for their very helpful comments. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 352015383 – SFB 1330 B4.

3.8 Appendix

3.8.1 Derivation of cross-power spectral density in opposingly-delayed noise

In ODN, two two-channel signals $u(t) = [u(t)\ u(t + ITD)]$ and $z(t) = [z(t)\ z(t - ITD)]$ with opposite ITDs, ITD and -ITD, are summed. The cross-power spectral density (CPSD) functions are

$$\begin{aligned} S_{UU}(\omega) &= 0.5e^{iITD\omega}, \\ S_{ZZ}(\omega) &= 0.5e^{-iITD\omega}. \end{aligned} \tag{3.10}$$

The power spectral density is 0.5 1/Hz each, so that the ODN has the same energy as the SDN. Summation of the time signals is equivalent to a summation of their CPSD functions, which leads to

$$S_{UZ} = S_{UU}(\omega) + S_{ZZ}(\omega) = \cos(\omega ITD). \tag{3.11}$$

This resulting cosine pattern is determined by the sum of the CPSDs' phases adding up or canceling each other at different frequencies. This normalized CPSD $C(\omega)$ represents the coherent energy of the signals as a function of frequency (Gardner, 1992),

$$C(\omega) = \frac{|S_{UZ}(\omega)|}{\sqrt{S_{UU}(\omega)S_{ZZ}(\omega)}} = |\cos(\omega ITD)|. \tag{3.12}$$

If $|\gamma(\tau)|$ is based on an ensemble average, then $C(\omega) = \mathcal{F}\{|\gamma(\tau)|\}$, with $\mathcal{F}\{\bullet\}$ the fourier transform. As a continuous function of ω it gives a coherence for any frequency ω representing an infinitesimally small bandwidth, illustrated as continuous lines in Fig. 3.1(D). The coherence for peripherally filtered, i.e. finite-bandwidth signals is an average of the frequencies' normalized CPSDs $C(\omega)$. The coherence decreases with increasing ITD and increasing bandwidth, as illustrated by the bars in Fig. 3.1(D). Two superimposed noises with ITD = ± 2 ms are in phase at 500 Hz. At 625 Hz, however, they have IPDs of $\pi/2$ and $-\pi/2$, respectively. The coherence between left and right signals at 625 Hz is therefore zero.

References

Bernstein, L. R., and Trahiotis, C. (**1995**). “Binaural interference effects measured with masking-level difference and with ITD- and IID-discrimination paradigms,” The Journal of the Acoustical Society of America **98**(1), 155–163, doi: `10.1121/1.414467`.

Bernstein, L. R., and Trahiotis, C. (**2014**). “Accounting for binaural detection as a function of masker interaural correlation: Effects of center frequency and bandwidth,” The Journal of the Acoustical Society of America **136**(6), 3211–3220, doi: `10.1121/1.4900830`.

Bernstein, L. R., and Trahiotis, C. (**2015**). “Converging measures of binaural detection yield estimates of precision of coding of interaural temporal disparities,” The Journal of the Acoustical Society of America **138**(5), EL474–EL479, doi: `10.1121/1.4935606`.

Bernstein, L. R., and Trahiotis, C. (**2017**). “Binaural detection-based estimates of precision of coding of interaural temporal disparities across center frequency,” The Journal of the Acoustical Society of America **141**(5), 3973–3973, doi: `10.1121/1.4989060`.

Bernstein, L. R., and Trahiotis, C. (**2018**). “Effects of interaural delay, center frequency, and no more than “slight” hearing loss on precision of binaural processing: Empirical data and quantitative modeling,” The Journal of the Acoustical Society of America **144**(1), 292–307, doi: `10.1121/1.5046515`.

Bernstein, L. R., and Trahiotis, C. (**2020**). “A crew of listeners with no more than “slight” hearing loss who exhibit binaural deficits also exhibit higher levels of stimulus-independent internal noise,” The Journal of the Acoustical Society of America **147**(5), 3188–3196, doi: `10.1121/10.0001207`.

Best, V., Gallun, F. J., Carlile, S., and Shinn-Cunningham, B. G. (**2007**). “Binaural interference and auditory grouping,” The Journal of the Acoustical Society of America **121**(2), 1070–1076, doi: `10.1121/1.2407738`.

Biberger, T., and Ewert, S. D. (**2016**). “Envelope and intensity based prediction of psychoacoustic masking and speech intelligibility,” The Journal of the Acoustical Society of America **140**(2), 1023–1038, doi: `10.1121/1.4960574`.

Biberger, T., and Ewert, S. D. (**2017**). "The role of short-time intensity and envelope power for speech intelligibility and psychoacoustic masking," The Journal of the Acoustical Society of America **142**(2), 1098–1111, doi: `10.1121/1.4999059`.

Bourbon, W. T., and Jeffress, L. A. (**1965**). "Effect of Bandwidth of Masking Noise on Detection of Homophasic and Antiphasic Tonal Signals," 3.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**1999**). "The contribution of static and dynamically varying ITDs and IIDs to binaural detection," The Journal of the Acoustical Society of America **106**(2), 979–992, doi: `10.1121/1.427110`.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**a). "Binaural processing model based on contralateral inhibition. I. Model structure," The Journal of the Acoustical Society of America **110**(2), 1074–1088, doi: `10.1121/1.1383297`.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**b). "Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters," The Journal of the Acoustical Society of America **110**(2), 1105–1117, doi: `10.1121/1.1383299`.

Dau, T., Piechowiak, T., and Ewert, S. D. (**2013**). "Modeling within- and across-channel processes in comodulation masking release," The Journal of the Acoustical Society of America **133**(1), 350–364, doi: `10.1121/1.4768882`.

Dietz, M., Encke, J., Bracklo, K. I., and Ewert, S. D. (**2021**). "Tone detection thresholds in interaurally delayed noise of different bandwidths," Acta Acustica **5**, 60, doi: `10.1051/aacus/2021054`.

Encke, J., and Dietz, M. (**2022**). "A hemispheric two-channel code accounts for binaural unmasking in humans," Communications Biology **5**(1), 1122, doi: `10.1038/s42003-022-04098-x`.

Gardner, W. A. (**1992**). "A unifying view of coherence in signal processing," Signal Processing **29**(2), 113–140, doi: `10.1016/0165-1684(92)90015-O`.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hearing Research **47**(1-2), 103–138, doi: `10.1016/0378-5955(90)90170-T`.

Green, D. M. (**1966**). "Signal-Detection Analysis of Equalization and Cancellation Model," The Journal of the Acoustical Society of America **40**(4), 833–838, doi: `10.1121/1.1910155`.

Hacker, M. J., and Ratcliff, R. (**1979**). "A revisted table of d′ for M-alternative forced choice," Perception & Psychophysics **26**(2), 168–170, doi: `10.3758/BF03208311`.

Hirsh, I. J. (**1948**). "The Influence of Interaural Phase on Interaural Summation and Inhibition," The Journal of the Acoustical Society of America **20**(4), 536–544, doi: `10.1121/1.1906407`.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a Gammatone filterbank," Acta Acustica united with Acustica **88**(3), 433–442.

Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). "Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments," The Journal of the Acoustical Society of America **104**(4), 2412–2425, doi: `10.1121/1.423773`.

Jeffress, L. A. (**1948**). "A place theory of sound localization.," Journal of Comparative and Physiological Psychology **41**(1), 35–39, doi: `10.1037/h0061495`.

Joris, P. X., de Sande, B. V., Louage, D. H., and van der Heijden, M. (**2006**). "Binaural and cochlear disparities," Proceedings of the National Academy of Sciences **103**(34), 12917–12922, doi: `10.1073/pnas.0601396103`.

Just, D., and Bamler, R. (**1994**). "Phase statistics of interferograms with applications to synthetic aperture radar," Applied Optics **33**(20), 4361, doi: `10.1364/AO.33.004361`.

Kolarik, A. J., and Culling, J. F. (**2010**). "Measurement of the binaural auditory filter using a detection task," The Journal of the Acoustical Society of America **127**(5), 3009–3017, doi: `10.1121/1.3365314`.

Langford, T. L., and Jeffress, L. A. (**1964**). "Effect of Noise Crosscorrelation on Binaural Signal Detection," The Journal of the Acoustical Society of America **36**(8), 1455–1458, doi: `10.1121/1.1919224`.

Leibold, C., and Grothe, B. (**2015**). "Sound localization with microsecond precision in mammals: What is it we do not understand?," e-Neuroforum **6**(1), 3–10, doi: `10.1007/s13295-015-0001-3`.

Lüddemann, H., Riedel, H., and Kollmeier, B. (**2007**). "Logarithmic Scaling of Interaural Cross Correlation: A Model Based on Evidence from Psychophysics and EEG," in *Hearing – From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey, Springer, Berlin, Heidelberg, pp. 379–388, doi: `10.1007/978-3-540-73009-5_41`.

Marquardt, T., and McAlpine, D. (**2009**). "Masking with interaurally "double-delayed" stimuli: The range of internal delays in the human brain," The Journal of the Acoustical Society of America **126**(6), EL177–EL182, doi: `10.1121/1.3253689`.

Mc Laughlin, M., Chabwine, J. N., van der Heijden, M., and Joris, P. X. (**2008**). "Comparison of Bandwidths in the Inferior Colliculus and the Auditory Nerve. II: Measurement Using a Temporally Manipulated Stimulus," Journal of Neurophysiology **100**(4), 2312–2327, doi: `10.1152/jn.90252.2008`.

McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). "A neural code for low-frequency sound localization in mammals," Nature Neuroscience **4**(4), 396–401, doi: `10.1038/86049`.

McFadden, D., and Pasanen, E. G. (**1976**). "Lateralization at high frequencies based on interaural time differences," The Journal of the Acoustical Society of America **59**(3), 634–639, doi: `10.1121/1.380913`.

McNemar, Q. (**1969**). *Psychological Statistics*, 4th ed ed. (Wiley, New York).

Oxenham, A. J., and Dau, T. (**2001**). "Modulation detection interference: Effects of concurrent and sequential streaming," The Journal of the Acoustical Society of America **110**(1), 402–408, doi: `10.1121/1.1373443`.

Piechowiak, T., Ewert, S. D., and Dau, T. (**2007**). "Modeling comodulation masking release using an equalization-cancellation mechanism," The Journal of the Acoustical Society of America **121**(4), 2111–2126, doi: `10.1121/1.2534227`.

Pollack, I., and Trittipoe, W. J. (**1959**). "Binaural Listening and Interaural Noise Cross Correlation," The Journal of the Acoustical Society of America **31**(9), 1250–1252, doi: `10.1121/1.1907852`.

Rabiner, L. R., Laurence, C. L., and Durlach, N. I. (**1966**). "Further Results on Binaural Unmasking and the EC Model," The Journal of the Acoustical Society of America **40**(1), 62–70, doi: `10.1121/1.1910065`.

Robinson, D. E., and Jeffress, L. A. (**1963**). "Effect of Varying the Interaural Noise Correlation on the Detectability of Tonal Signals," The Journal of the Acoustical Society of America **35**(12), 1947–1952, doi: `10.1121/1.1918864`.

Sondhi, M. M., and Guttman, N. (**1966**). "Width of the Spectrum Effective in the Binaural Release of Masking," The Journal of the Acoustical Society of America **40**(3), 600–606, doi: `10.1121/1.1910124`.

Stern, R. M., and Colburn, H. S. (**1978**). "Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position," The Journal of the Acoustical Society of America **64**(1), 127–140, doi: `10.1121/1.381978`.

Stern, R. M., Colburn, H. S., Bernstein, L. R., and Trahiotis, C. (**2019**). "The fMRI Data of Thompson et al. (2006) Do Not Constrain How the Human Midbrain Represents Interaural Time Delay," Journal of the Association for Research in Otolaryngology **20**(4), 305–311, doi: `10.1007/s10162-019-00715-5`.

Thompson, S. K., von Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T. D., and McAlpine, D. (**2006**). "Representation of interaural time delay in the human auditory midbrain," Nature Neuroscience **9**(9), 1096–1098, doi: `10.1038/nn1755`.

van de Par, S., and Kohlrausch, A. (**1999**). "Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters," The Journal of the Acoustical Society of America **106**(4), 1940–1947, doi: `10.1121/1.427942`.

van der Heijden, M., and Trahiotis, C. (**1998**). "Binaural detection as a function of interaural correlation and bandwidth of masking noise: Implications for estimates of spectral resolution," The Journal of the Acoustical Society of America **103**(3), 1609–1614, doi: `10.1121/1.421295`.

van der Heijden, M., and Trahiotis, C. (**1999**). "Masking with interaurally delayed stimuli: The use of "internal" delays in binaural detection," The Journal of the Acoustical Society of America **105**(1), 388–399, doi: `10.1121/1.424628`.

Verhey, J. L., and van de Par, S. (**2020**). “Binaural frequency selectivity in humans,” European Journal of Neuroscience **51**(5), 1179–1190, doi: `10.1111/ejn.13837`.

Yost, W. A., and Sheft, S. (**1989**). “Across-critical-band processing of amplitude-modulated tones,” The Journal of the Acoustical Society of America **85**(2), 848–857, doi: `10.1121/1.397556`.

CHAPTER 4

Fast binaural processing but sluggish masker representation reconfiguration

This chapter is a formatted reprint of
Eurich, B. and Dietz, M. (2023). Fast binaural processing but sluggish masker representation reconfiguration. *The Journal of the Acoustical Society of America*, 154(3), 1862–1870, doi: `10.1121/10.0021072`.

Author contributions: BE reviewed the literature and background, developed the concept and implemented the model, computed the predictions, prepared the figures and wrote as well as revised the manuscript. MD participated in developing the concept as well as in improving and revising the manuscript.

4.1 Abstract

Perceptual organization of complex acoustic scenes requires fast binaural processing for accurate localization or lateralization based on short single-source-dominated glimpses. This sensitivity also manifests in the ability to detect rapid oscillating interaural time and phase differences as well as interaural correlation. However, binaural processing has also been termed “sluggish” based on experiments that require binaural detection in a masker with an additional binaural cue change in temporal proximity. The present study shows that the temporal integration windows obtained from data on binaural sluggishness cannot account for the detection of rapid binaural oscillations. A model with fast IPD encoding but a slower process of updating the internal representation of the masker IPD statistics accounted fo both, experiments of the “fast” and the “sluggish”‘ categories.

4.2 Introduction

The extraordinary precision of encoding interaural phase difference (IPD) allows for binaural unmasking and low-frequency sound localization. Psychophysical characterization of binaural unmasking and sound localization, however, led to different interpretations concerning the processing speed of the binaural system. On the one hand, subjects are able to lateralize and detect a change in ITD as brief as 3 - 6 ms (Reed *et al.*, 2016). Also, detection thresholds of broadband binaural beat or “Phasewarp” stimuli, (Siveke *et al.*, 2008) or of oscillating interaural correlation (“Oscor” stimulus, Grantham, 1982; Grantham and Wightman, 1979; Gatehouse and Akeroyd, 2006; Siveke *et al.*, 2008) embedded in uncorrelated noise have been measured up to beat- or oscillation frequencies of 1024 Hz and 128 Hz, respectively (Siveke *et al.*, 2008). All these are indicators of very fast binaural processing. Modeling suggested that the temporal resolution is primarily limited only by the bandwidth of the auditory filters (Dietz *et al.*, 2008), i.e. their ring time. On the other hand, many tone-in-noise detection experiments have led to the interpretation of a “sluggish”‘ binaural system. For example, Kollmeier and Gilkey (1990) found elevated detection thresholds for an antiphasic tone (S_π) in temporal proximity of an inversion of the masker’s IPD. When the masker changed from antiphasic N_π, which gives no binaural benefit, to N_0, thresholds drop only gradually as the target tone burst is moved away from the transition moment. For a full binaural unmasking over

100 ms separation is required, much more than in monaural forward masking experiments. The data were well fitted by a double-sided exponential integration window with equivalent rectangular durations (ERD) in the range $33.2 \leq \text{ERD} \leq 83.2$ ms. When replacing the IPD inversion by a diotic 15 dB masker attenuation, integration times of only $11.9 \leq \text{ERD} \leq 26.0$ ms were fitted. Binaural sluggishness was also supported by Grantham and Wightman (1979). They used the above-mentioned oscillating interaural correlation "Oscor" noise as a masker and presented a short interaurally out of phase (S_π) tone pip coinciding with a moment where the instantaneous masker correlation is either 1 or -1, i.e. where the masker briefly resembled N_0 or N_π, respectively. The binaural masking level difference (BMLD) disappared already at a modulation frequency of 4 Hz. Grantham and Wightman (1979) explained the effect by assuming different monaural and binaural temporal analysis windows with binaural integration times of $44 \leq \text{ERD} \leq 140$ ms. However, averaging interaural cues over such a long time window is likely to conflict with the "fast" studies mentioned above. To our knowledge there is no model that can account for both classes of experimental data without significantly changing the integration time constant.

We hypothesize interaural differences to be encoded with a high temporal resolution primarily limited by peripheral filter ring times. Furthermore, we follow the hypothesis of Yost (1985), who suggested that binaural unmasking relies on an estimate of the masker parameters. We hypothesize that if a task requires re-estimation of masker parameters, this higher-stage operation is the cause of the sluggish behavior. The two possible cases can be exemplified with the Oscor stimulus: As the oscillation frequency increases, the changes in perceived masker width move closer in time to the target tone, interfering with detection by reducing the contrast in perceived width between target and masker. This means the widening cue induced by the antiphasic target tone is disrupted by the widening and narrowing induced by the correlation oscillations of the masker. The instantaneous correlation of the masker at the moment the target is added cannot be exploited by human listeners (Grantham and Wightman, 1979). In contrast, if the task is to detect the presence of the oscillating correlation in a static masker, the estimation process is not needed. In this case, the auditory system can exploit the few millisecond short moments of high correlation (Siveke *et al.*, 2008). The rapid sensory encoding of IPD allows these rapid modulations to be perceived as a fluttering or intra-head rotation pattern (Witton *et al.*, 2000), which is the primary cue for such a task.

In this work we first show that the temporal integration windows obtained from data on binaural detection close in time to a masker IPD change (sluggish category) cannot account for the detection of rapid correlation oscillations. We then suggest a model with fast IPD encoding but a slow reformation of the internal representation of the masker. To simulate reduced sensitivity in temporal proximity to masker IPD changes, the model compares the stimulus internal representation to a template. As this template, a low-pass filtered internal representation of the masker is used. The low pass resembles the sluggish reorganization of the internal masker representation. Without changes to the model or its temporal parameters, it can account for both, data that requires fast and data that previously required sluggish binaural processing.

4.3 Simulations

4.3.1 Temporal integration does not account for oscillating correlation detection

We exemplarily tested whether the integration windows derived from studies that fall into the sluggish category can still be compatible with detection of rapid oscillations of interaural correlation (Siveke *et al.*, 2008). The integrations times fitted by Kollmeier and Gilkey (1990) ($33.2 \leq \text{ERD} \leq 83.2\,\text{ms}$) are similar to Holube *et al.* (1998) in their comparable step condition ($40 \leq \text{ERD} \leq 69\,\text{ms}$). Replacing the masker correlation step with a cosine correlation oscillation, as in the cosine condition of Holube *et al.* (1998) and the similar experiment of Grantham and Wightman (1979), resulted in longer windows of $91 \leq \text{ERD} \leq 122\,\text{ms}$ and $44 \leq \text{ERD} \leq 140\,\text{ms}$, respectively.

The peak-to-peak interaural correlation of an unmasked, gammatone-filtered Oscor stimulus can be reduced by temporal integration at the output of the binaural interaction. We compared the peak-to-peak correlation after temporal integration to the just-noticeable difference (JND) from zero correlation which is 0.3 (Boehnke *et al.*, 2002). We used an oscillation frequency of 64 Hz being the highest of the modulation frequencies measured by Siveke *et al.* (2008) that can be preserved by a gammatone filter centered at 500 Hz. If the peak-to-peak modulation of interaural correlation of the filtered Oscor stimulus exceeds the JND of 0.3, we expect that the corresponding time constant allows for the detection of oscillating correlation as reported by Siveke *et al.* (2008).

Fig. 4.1(A) shows the interaural correlation resulting from gammatone-filtering and temporal integration using a double-sided exponential integration window with ERDs in the range reported in the literature. While the correlation modulation in the stimulus oscillates between -1.0 and +1.0, it is already reduced to the range $[-0.5 \;\; 0.5]$ by gammatone filtering. Fig. 4.1(B) shows that any applied temporal integration with ERD $> 10\,\mathrm{ms}$ reduced the peak-to-peak correlation below the estimated JND (dotted line). Integration windows of, e.g., ERD $= 30\,\mathrm{ms}$ (i.e. $\tau = 15\,\mathrm{ms}$ for a double-sided exponential window) are already below the lower boundary of all models that fall into the sluggish category but at the same time they are far too large to preserve any perceivable correlation from the Oscor. For comparison: The double-sided exponential window used in the comprehensive and widely used model of Breebaart *et al.* (2001) had an ERD of 60 ms. Thus, we conclude that detection of rapid correlation modulation cannot be explained by the integration time constants that have been widely attributed to binaural detection and we cannot find any “compromise integration window” satisfying both categories.

4.3.2 A Model for Both Fast and Sluggish Processing

Having observed that temporal integration of interaural correlation alone cannot explain results on detecting rapid modulations and at the same time sluggishness, we developed a model with the goal to account for the aforementioned categories at the same time. It was designed to reproduce listeners’ behavior expected from the stimulus as processed according to our hypothesis: Interaural differences are available for detection with the temporal resolution only limited by basilar-membrane processing, at least at frequencies below about 1000 Hz. Binaural sluggishness accordingly appears if a change in masker IPD statistics, entailing reformation of its internal representation, interferes with target-in-masker detection. In this model, the internal representation includes the instantaneous interaural correlation, coherence, and phase.

Peripheral Processing

The left and right input signals were processed with a fourth-order gammatone filter. This represents basilar-membrane bandpass filtering. The filterbank imple-

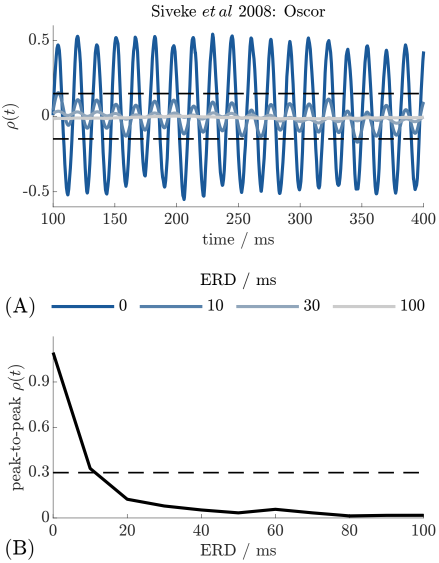

Figure 4.1: **(A)** Oscor stimulus where the interaural correlation is modulated with 64 Hz after gammatone filtering and temporal integration using a double-sided exponential window with different ERDs. **(B)** Resulting peak-to-peak correlation for the Oscor stimulus as shown in (A). The dashed lines at ±0.15 in (A) and +0.3 in (B) denote the assumed JND to detect the modulation.

mentation by Hohmann (2002) was used. The tone-in-noise experiments that we exemplarily chose for the present single-channel simulations (Kollmeier and Gilkey, 1990; Grantham and Wightman, 1979; Buss and Hall III, 2011) determined the used filter center frequency to be 500 Hz. The filter bandwidth (ERB $=$ 79 Hz) or its corresponding ring time (ERD $=$ 6.3 ms) limits the speed with which interaural cues can change, also the experiments by Siveke *et al.* (2008) are simulated only at 500 Hz, to keep a fair comparison that is not influenced by employing higher frequency channels and thus filters with a shorter ring time.

Binaural Processing

The instantaneous complex correlation coefficient $\gamma(t)$ serves as the internal representation. It was derived from the analytical (i.e. complex-valued) left and right signals $l(t)$ and $r(t)$ provided by the gammatone filter:

$$\gamma(t) =< \frac{l^*(t)r(t)}{\sqrt{|l(t)|^2|r(t)|^2}} >, \tag{4.1}$$

where $\bullet^*$ denotes the complex conjugate and $< \bullet >$ the average over an ensemble of stimulus tokens[1]. $\gamma(t)$ was used because it conveniently combines information about the instantaneous interaural correlation $\rho(t) = \Re\{\gamma(t)\}$, instantaneous coherence $|\gamma(t)|$ and instantaneous IPD $\arg\{\gamma(t)\}$ (Encke and Dietz, 2022). The resulting binaural display represents the information available to the detector.

The hypothesis of a gradual reformation of the masker internal representation resulted from the observation that binaural detection is impaired in presence of an interaurally modulated *masker* but not for detection of a modulated *target*. We modeled the gradual reformation of the internal representation of the masker by employing a parallel path in which $\gamma(t)$ of a masker-alone version of the stimulus involved low-pass filtering. A double-sided exponential window with a time constant of $\tau = 30\,\mathrm{ms}$ for each lobe was used. The ERD was thus $2\tau = 2 \times 30\,\mathrm{ms} = 60\,\mathrm{ms}$, oriented on the window fits by Kollmeier and Gilkey (1990) and on the ERD $=$ 60 ms used in the model of Breebaart *et al.* (2001).

$$\gamma_{M,\,\mathrm{lp}}(t) =< \mathrm{lp}[(\frac{l^*(t)r(t)}{\sqrt{|l(t)|^2|r(t)|^2}}] >, \tag{4.2}$$

Subsequent to the reformation-mimicking low-pass filtering, the average over an ensemble of 200 stimulus tokens was used in Eq. 4.1 and Eq. 4.2. This describes the expected behavior of a group of listeners after many repetitions of the experiment. As the results of Kollmeier and Gilkey (1990) reveal comparable binaural forward and backward masking effects, a zero-phase forward-reverse filter was used. The internal representation of the masker $\gamma_{M,\,\mathrm{lp}}(t)$ is then contrasted with the internal representation of the complete stimulus $\gamma(t)$, i.e. masker plus target without low-pass filtering. The real parts of $\gamma(t)$ and $\gamma_{M,\,\mathrm{lp}}(t)$, i.e. interaural correlation, are plotted in Fig. 4.2. For stimuli with changing masker correlation, such as Kollmeier and Gilkey (1990), the difference between template and target is therefore reduced

---

[1] For consistent processing in the two internal representations used in the model, the average over an ensemble of stimulus tokens was taken for numerator and denominator together instead of, as used in Encke and Dietz (2022), separately. This is irrelevant for the present simulations as it only negotiates differences between the left and right signals' amplitudes.

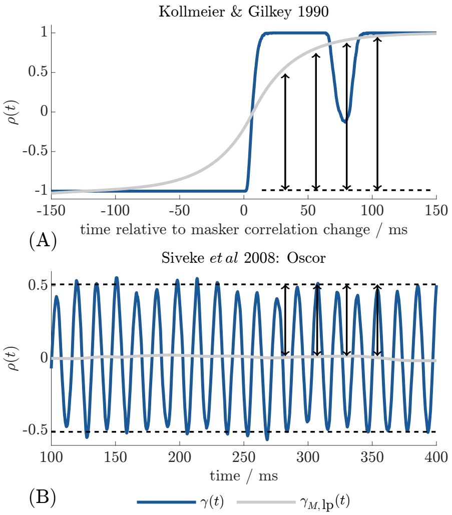

Figure 4.2: Real part (i.e. interaural correlation $\rho(t)$) of the stimulus internal representation $\gamma(t)$ and of the masker internal representation $\gamma_{M,\,\mathrm{lp}}(t)$, as used in the present model. Arrows symbolize the maximum possible binaural cue. **(A)** stimulus as used in Kollmeier and Gilkey (1990) with 80 ms delay time between the masker changing from N_π to N_0 and the S_π target; SNR = -16 dB. **(B)** Oscor stimulus as used in Siveke *et al.* (2008), Oscillation frequency 64 Hz, SNR = 4 dB.

in temporal proximity to the masker correlation change, reducing the detection cue. An oscillating correlation or Phasewarp (Siveke *et al.*, 2008), however, is preserved as the required temporal precision is still available to the detector from the unfiltered target $\gamma(t)$. The low-pass filtering prior to the ensemble averaging in the internal representation of the masker represents the hypothesis that a comparatively slow adaptation to a change in the $\gamma(t)$ statistics of the masker impairs detection. This means, the binaural feature allows both detection of an oscillating pattern and detection of a change in perceived width with the same backend.

As in Encke and Dietz (2022) and Eurich *et al.* (2022), the unity-limited interaural representations $\gamma(t)$ and $\gamma_{M,\mathrm{lp}}(t)$ are Fisher z (i.e. inverse hyperbolic tangent) transformed for the purpose of variance normalization (McNemar, 1969; Just and Bamler, 1994), as often applied in psychophysics (e.g., Lüddemann *et al.*, 2007; Bernstein and Trahiotis, 2017). $\gamma(t)$ and $\gamma_{M,\mathrm{lp}}(t)$ are multiplied by a model parameter $\hat{\rho} < 1$ (Encke and Dietz, 2022; Eurich *et al.*, 2022) to avoid an infinite sensitivity to deviations from a coherence of unity. This is equivalent to adding uncorrelated noise to the two input signals. The inner representation of the binaural pathway is thus:

$$\zeta(t) = z[\hat{\rho}\gamma(t)]; \tag{4.3}$$

$$\zeta_{M,\mathrm{lp}}(t) = z[\hat{\rho}\gamma_{M,\mathrm{lp}}(t)]; \tag{4.4}$$

where $z[\bullet]$ is the Fisher z-transform applied to the modulus of $\gamma(t)$ while leaving the argument unchanged.

The model extracts the instantaneous absolute difference between the internal representations of the stimulus and that of the masker:

$$b(t) = |\zeta(t) - \zeta_{M,\mathrm{lp}}(t)|. \tag{4.5}$$

As the model was required to detect ongoing temporal patterns, a sequence of independent observations was extracted, based on the multiple-looks principle (Viemeister and Wakefield, 1989). These observations are combined in an optimal manner:

$$b_{\mathrm{opt}} = \sqrt{\sum_t b(t)^2}. \tag{4.6}$$

The model parameter σ_{bin} represents the standard deviation of an internal noise with a Gaussian distribution of amplitudes. It adjusts the sensitivity of the binaural path in order to model the sensitivity index d'_{bin} of discriminating between the noise alone (N) interval and the signal plus noise ($S+N$) interval:

$$d'_{\mathrm{bin}} = \frac{b_{\mathrm{opt},\,S+N} - b_{\mathrm{opt}\,N}}{\sigma_{\mathrm{bin}}}. \tag{4.7}$$

The optimal combination in a multiple-looks manner (equation 4.6) enables the model to inherently code higher sensitivity for longer detection cues.

Monaural Processing

In order to simulate experiments that compare binaural and monaural processing speed, we also designed a simple monaural path. It is sensitive to changes in stimulus instantaneous envelope power $P(t)$. The envelope is the modulus of the complex-valued filter output. For the stimuli employed in this study, it is sufficient to evaluate only one side. Similar to the binaural path, the temporal resolution of $P(t)$ is only limited by the ring time of the gammatone filter (ERD $= 6.3\,\mathrm{ms}$). The model considers the difference in ensemble-average envelope power between the signal-plus-noise and noise-alone intervals:

$$\Delta P(t) =< P_{N+S}(t) > - < P_N(t) > \tag{4.8}$$

As in the binaural pathway, the instantaneous power differences $\Delta P(t)$ are treated as independent observations and thus combined in an optimal manner:

$$\Delta P_{\mathrm{opt}} = \sqrt{\sum_t [\frac{\Delta P(t)}{< P_N(t) >}]^2} \tag{4.9}$$

The processing accuracy is limited by a level-dependent internal noise with a Gaussian distribution of amplitudes and a standard deviation of σ_{mon} (Eurich *et al.*, 2022):

$$d'_{\mathrm{mon}} = \frac{\Delta P_{\mathrm{opt}}}{\sigma_{\mathrm{mon}}}. \tag{4.10}$$

4.3.3 Quantitative Predictions of Detection Thresholds

We used the described model paths to estimate detection thresholds across different experiments. The sensitivity indices d' of monaural and binaural path are therefore combined to the overall sensitivity index, assuming independent information channels (Green and Swets, 1966):

$$d'_{\mathrm{bin+mon}} = \sqrt{d'^2_{\mathrm{bin}} + d'^2_{\mathrm{mon}}} \tag{4.11}$$

The d' that corresponds to the experiment-specific detection thresholds was obtained via table-lookup [numerical evaluation in (Hacker and Ratcliff, 1979)]. This depends on the number of intervals as well as the specific staircase procedure used in the simulated experiments. As in Eurich *et al.* (2022), for each condition the model was evaluated for a range of target levels. This provided the psychometric function. The predicted detection threshold was obtained from a linear fit to the logarithmic d'. The model parameters σ_{bin} and σ_{mon} were manually adjusted to optimize the prediction accuracy. The resulting parameter values are given in Table 4.1. The aim was to see whether a single detector with a single time constant can at the

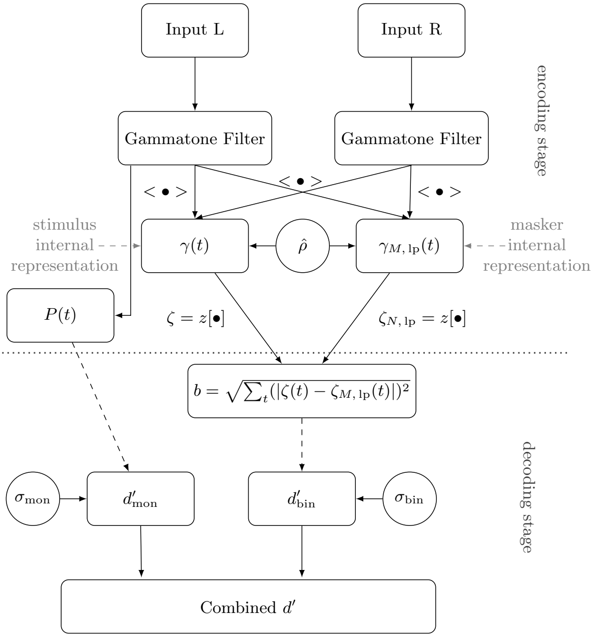

Figure 4.3: Flowchart of the model as used to quantitatively predict binaural detection thresholds. Dashed arrows symbolize that the sensitivity indices d' are obtained by comparing signal-plus-noise with noise-alone.

same time account for different paradigms supporting fast and sluggish processing, rather than to maximize precision concerning a certain paradigm. The following experiments were chosen as a representative selection concerning evidence for fast and sluggish binaural processing. All experiments can be simulated based on the output of a single bandpass filter centered at 500 Hz. This ensures that IPDs in the temporal fine structure are the dominant cue for detection and allows for a good comparison. Figure 4.4 shows experimental thresholds and predictions with continuous lines denoting predictions for conditions with partly different IPDs of masker and target tone, dotted lines conditions where masker and tone always have the same IPD.

Due to the good agreement between simulations and the critical trends in the data, the experiments are described together with the predictions.

Experiment	category	Target	$\hat{\rho}$	σ_{bin}	σ_{mon}	ERD / ms
Kollmeier and Gilkey (1990)	sluggish	S_π	0.9	12	500	60
Grantham and Wightman (1979)	sluggish	S_π	0.9	20	350	60
Siveke *et al.* (2008)	fast	Oscor, Phasewarp	0.9	22	200	60
Dietz *et al.* (2008)	fast	Phasewarp	0.9	22	200	60
Buss and Hall III (2011)	fast	S_0, S_π	0.9	20	200	60

Table 4.1: Summary of the simulated experiments and predictions. *Columns 1 - 3*: Simulated experiment, rough category of binaural processing speed concluded by the authors, target signal. *Columns 4 - 7*: Used model parameters: $\hat{\rho} < 1$: Maximum coherence (internal noise); σ_{bin}: Standard deviation of the internal noise to determine the absolute performance of the binaural path; σ_{mon}: Standard deviation of the level-dependent internal noise to determine the acuity of the monaural pathway; equivalent rectangular duration of the temporal window used for the template.

Kollmeier & Gilkey, 1990

A 20-ms 500 Hz S_π tone is to be detected in a 750-ms noise masker whose interaural correlation changes from 1 to -1, or vice versa, after 375 ms. This means the tone interaural correlation (-1) differs from the masker interaural correlation only when appearing in either the first or the second half of the masker. Tone detection thresholds were reported as a function of the delay time of the tone relative to the moment of masker correlation change. The 0 dB point corresponds to each subject's and the model's detection thresholds for a reference $N_\pi S_\pi$ condition with a static masker, respectively. Experimental thresholds and simulations are shown in Fig. 4.4 (A). Thresholds are highest when the S_π tone is presented during the N_π segment and are gradually getting lower as the tone is moved to the N_0 part of the masker. Quantitatively, however, considerable difference are apparent across subjects. The predictions are generally within the range of the data. In another condition, instead of providing an interaural cue, one half of the masker was attenuated by 15 dB. The fact that thresholds were elevated up to about 120 ms into the N_0 part but now only about 20 ms into the attenuated part has in the past been interpreted as a longer binaural than monaural temporal analysis window or sluggish binaural processing. This considerably steeper threshold decrease is clearly reproduced by the model. Subjects sometimes performed worse in the conditions with the IPD or SNR transitions in the masker, albeit up to 200 ms apart, than in the $N_\pi S_\pi$ reference condition with a static masker, as indicated by thresholds above 0 dB. The model, however, predicts the same thresholds in both cases, as expected. The fact that thresholds rise shortly before the masker level transition in Fig. 4.4 (A), lower left panel, is so-called backward masking. Because the present model does not include the cortical effects associated with backward masking (Zwicker and Fastl, 1999), thresholds do not increase for tones appearing before the masker level increase. This leads to a slight offset between the data and the predictions, while the slope is reproduced, being much steeper than in the conditions with a masker IPD switch.

Grantham & Wightman, 1979

Fig. 4.4 (B) shows data and simulations for a second example on binaural sluggishness. As pointed out in the introduction, the interaural correlation of the *masker* oscillates between -1 and +1 (Oscor). The target tone – a gated 17-ms 500-Hz S_π tone pip – was presented either at moments when the interaural correlation was 1 or -1. If the S_π tone coincides with the correlated, i.e. momentarily N_0-like masker, binaural unmasking is only evident at the very slowest oscillation frequencies of 0.5 or 1 Hz. At 4 Hz there is already virtually no binaural benefit anymore, which has been interpreted as evidence for binaural sluggishness. All features in the data are captured by the model. Predictions are within the range of between-subject deviations.

Siveke et al., 2008

In this study, rapidly oscillating interaural cues serve as *target*, providing evidence for fast binaural processing, with experimental results and simulations plotted in Fig. 4.4 (D). Wideband noise was generated that had fast changing interaural cues. Two types of interaural cue changes were employed: (1) Oscillating interaural correlation between -1 and +1 (Oscor); (2) a binaural beat, i.e. a linearly increasing and phase-wrapping IPD (Phasewarp). Both types of interaurally modulated noise could be detected in interaurally uncorrelated noise up to modulation frequencies of 128 Hz and 1024 Hz in case of the Phasewarp. The oscillation- or beat-rate limit is thus more than one order of magnitude higher than in the above-mentioned data by Grantham and Wightman (1979) where the Oscor is employed as a *masker*. Our model replicates at the same time sluggish behavior [Oscor as masker, Grantham and Wightman (1979)] and non-sluggish behavior [Oscor/Phasewarp as target, Siveke *et al.* (2008)]. The sensitivity decline to detect very rapidly oscillating correlation or phase (64 Hz) is due to the auditory filter bandwidth. The interaurally correlated frequency components are separated by the modulation frequency. Thus, the corresponding left-right filter pairs see increasingly uncorrelated components. Increasing the center frequency of our analysis filter will allow detecting even higher phasewarp frequencies, as reported by Siveke *et al.* (2008). However, the modulation frequencies encoded by the 500 Hz filter (i.e. up to 64 Hz) are sufficient in order to account for the category of fast binaural processing, as opposite to sluggish processing. The fact that Dietz *et al.* (2008) obtained similar thresholds for a phase-warp bandlimitied to 0...550 Hz, we assume that for this type of sensitivity always arises from within-channel cues, rendering the simplistic single-channel model sufficient for this task and most comparable to the other tasks.

Buss & Hall, 2011

In contrast to Kollmeier and Gilkey (1990), the tone-in-noise detection data by Buss and Hall III (2011) suggest comparably fast binaural and monaural processing [Fig. 4.4 (C)]. The difference is that now the masker is attenuated for various durations but, in contrast to Kollmeier and Gilkey (1990) and Grantham and Wightman (1979), there is no change in the interaural configuration. N_0S_0 and N_0S_π thresholds are presented for a 20 ms long, ramped 500 Hz tone in wideband noise as a function of the duration of the 20 dB masker attenuation. The tone was either centered in or 20 ms shifted from the center of the attenuated part. The decreases in thresholds towards higher signal/masker intervals are considerably steeper than in the conditions in Kollmeier and Gilkey (1990) involving the masker-IPD transition. Furthermore, in contrast to Kollmeier and Gilkey (1990), thresholds decrease equally steep for both diotic and dichotic conditions. The model replicates these two core features of this experiment. However, predictions show a somewhat harder knee-point in thresholds than the data. This effect is associated with neural rate adaptation which was not employed in the present model in order to focus on the temporal processing hypothesis.

4.4 Discussion

Binaural processing has been shown to be similarly fast as monaural processing in detection of interaural cue modulations (Siveke *et al.*, 2008; Dietz *et al.*, 2008),

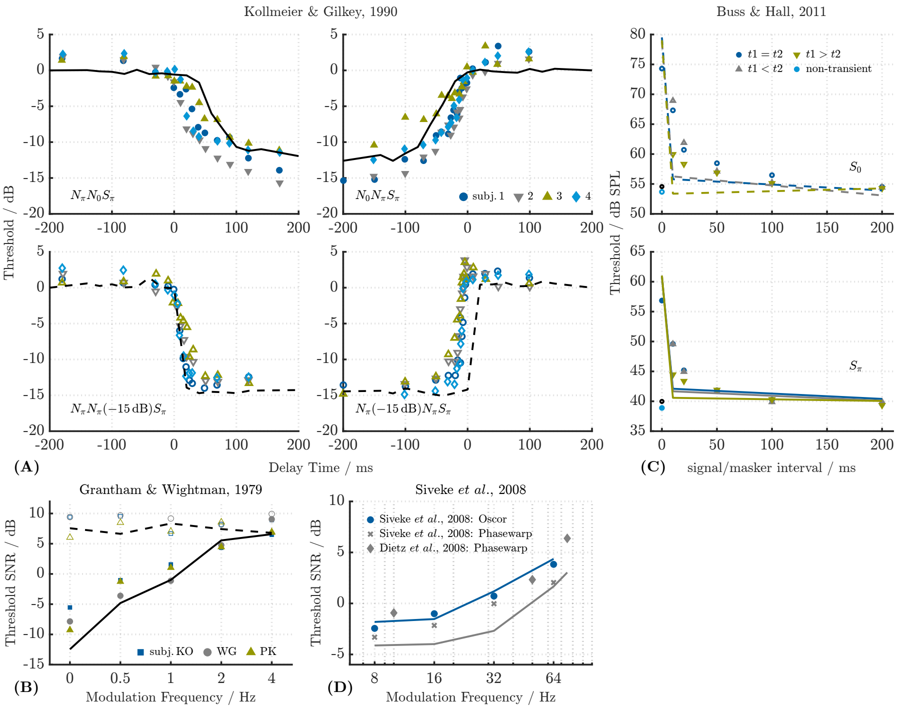

Figure 4.4: Simulation results for four exemplary data sets on sluggish and fast binaural processing. Symbols denote the original experimental thresholds with filled symbols for the conditions on binaural detection and open symbols for those on monaural detection. Continuous lines denote model predictions for binaural, dotted lines for monaural detection conditions. See main text for experimental details.

tone-in-noise detection (Buss and Hall III, 2011; Bischof *et al.*, 2023), ITD and ILD detection (Akeroyd and Bernstein, 2001), and alternating-ITD lateralization (Reed *et al.*, 2016) tasks. On the other hand, it has been characterized "sluggish" in other detection (Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Kolarik and Culling, 2009; Culling and Summerfield, 1998; Holube *et al.*, 1998), ITD discrimination (Kolarik and Culling, 2009), and speech intelligibility (Hauth and Brand, 2018; Culling and Mansell, 2013) tasks. We confirm previous statements that different tasks involve different aspects of binaural temporal processing (Akeroyd and Bernstein, 2001) – there is no compromise integration time accounting for both categories at the same time. With this result being very clear, such a compromise integration window is also not expected from other window shapes, such as a combination of a strongly weighted shorter and less weighted longer window as used by Bernstein *et al.* (2001). Instead, we accounted for a set of experiments from both categories with a model that compares an interval's internal representation to a low-pass filtered internal representation of the masker. This is a simplistic implementation of our hypothesis: Binaural sluggishness appears when detecting a static target in an interaurally modulated masker but not when detecting a modulated target in a static masker. This is a result of a slow reformation of the masker internal representation, or, in other words, an adaptation of the interaural masker profile estimate (Yost, 1985).

Such reformation takes effect in experiments as performed by Kollmeier and Gilkey (1990). There, the masker IPD is inverted (correlation is changed from -1 to 1), causing a change in its perceived spatiality (i.e. width). This interferes with the target cue which is also a change in perceived spatiality, because the S_π target causes IPD fluctuations when added to a diotic masker. Similarly, when adding an S_π target at the positive peak of a masker correlation oscillation (Grantham and Wightman, 1979), the continuously changing perceived masker spatiality interferes with target-induced widening. Hauth and Brand (2018) measured speech reception thresholds for spoken sentences in the presence of a masking noise with modulated IPD. Binaural unmasking was found to decay for modulation frequencies up to 4 Hz, similar to the above-mentioned tone-in-noise experiment by Grantham and Wightman (1979). We hypothesize the same across-time interference of the masker spatiality to limit binaural unmasking of speech in the presence of an interaurally modulated masker.

In contrast, to detect the correlation modulation as a target in uncorrelated noise (Siveke *et al.*, 2008), the perceived fluttering is sufficient. Thus, fast modulations in correlation can be detected while the accompanying width changes of an additional tone cannot (Witton *et al.*, 2000; Singh and Bharadwaj, 2021). In Buss and Hall III (2011) no sluggishness was observed although the dichotic condition requires interaural cue detection. Their masker, however, contained only an attenuation, no phase or correlation change and therefore no change in masker IPD statistics. The presented model preserves fast temporal fluctuations because temporal integration only affects the IPD template of the internal representation of the masker. Thus the model accounts for all mentioned stimulus categories.

Temporal-analysis-window lengths fitted to experimental results indicating sluggishness (e.g., Kollmeier and Gilkey, 1990; Holube *et al.*, 1998; Grantham and Wightman, 1979) extend over a wide range, both across studies but also across subjects within a study. This indicates that binaural sluggishness is not a fixed, inevitable phenomenon determined by the statistics of the interaural cues or by a hard-wired temporal integration but rather a consequence of an individual's decoding or interpretation of the encoded cues. According to Yasin and Henning (2012) such is consistent with McFadden (1966), Robinson and Trahiotis (1972) and Yost (1985), reasoning that the binaural system detects a target best when having an accurate estimate of the "masker profile"'. We add that this depends on the masker IPD statistics, i.e. on $\gamma_{M,\,\mathrm{lp}}(t)$, but not on its power profile. Therefore, our model based on IPD statistics alone appears to account for both sluggish as well as fast binaural processing data.

Simulating the four different paradigms with the same model supports our hypothesis that binaural processing speed, similar to monaural processing speed, is only limited by peripheral processing. Apparent sluggishness results from higher-level analysis of masker IPD statistics. In perceptual terms, this means that sluggishness occurs whenever a task requires the listener to adapt to a change in perceived width in order to be receptive to the target. According to Singh and Bharadwaj (2021), this is no longer possible for modulations above about 10 Hz. They further pointed out that monaural and binaural modulation share the same transitions to a flutter for modulation frequencies above about 7 to 10 Hz. This corresponds to periods of 100 to 143 ms which matches well with the duration of threshold convergence observed by, e.g., Kollmeier and Gilkey (1990). As Ross *et al.* (2014) concluded from neuromagnetic responses to binaural beats that perception of moving sounds are limited by the cortical rate of object formation, and as modulations below about 7 Hz have been classified as crucial for object binding (Shinn-Cunningham *et al.*, 2017), we associate the origin of binaural sluggishness with object formation.

However, similar to the detection of Oscor or Phasewarp, the lateralization of a noise burst with an ITD alternating rapidly with diotic noise segments (Reed *et al.*, 2016) does not result in a sluggish integration of IPD, because the two interleaved streams form spatially distinct objects. Instead of, as previous models, slowly averaging instantateous IPD on the sensory level, the proposed concept allows to collect binaural information on fast sensory data and puts any slow adaptation on the already separated objects, as suggested by Yabe *et al.* (2001). The concept is in line with typical duration dependence phenomena as characterized by Hafter *et al.* (1979), Hafter and Dye (1983), Houtgast and Plomp (1968), and Stecker (2014) including the knee point in N_0S_π thresholds as a function of signal duration at about 200 ms (Wilson and Fugleberg, 1987). It further agrees with the threshold decay measured by Kolarik and Culling (2009) for discriminating between two correlated or uncorrelated noise intervals, both containing a delayed noise of different length, differing only in the sign of the ITD. For short probe bursts, higher thresholds were observed compared to a condition where one interval contained only diotic noise and thus differed in coherence. The additional coherence cue is thought to cause the difference in thresholds. However, the similar timescales involved in object formation and integration suggest that both mechanisms contribute to the stable perception of an auditory object. In future works, the proposed model can be extended by subsequent temporal integration to account for saturating sensitivity at longer signals.

The proposed model captured the supposedly contradictory characteristics of the data, although different tasks that led to different conclusions on the binaural processing speed were simulated with the same model using the same temporal processing parameters. Only the internal-noise parameters σ_{bin} and σ_{mon} were changed between different tasks. The double-sided exponential window with its fixed ERD of $2\tau = 60$ ms is similar to the ERDs obtained by Boehnke *et al.* (2002) for detecting dynamic changes in interaural correlation and those obtained by Kollmeier and Gilkey (1990), each by assuming temporal integration of interaural correlation. Although Grantham and Wightman (1979) fitted substantially larger ERDs than Kollmeier and Gilkey (1990), their data are also well explained by the present model and $2\tau = 60$ ms. Simulations include the relationship of conditions on binaural and monaural detection, which differ in the “sluggishness data” [Fig. 4.4 (A) and (B)] but not in the “fast processing data” (Fig. 4.4(C) and (D)). Although the model captures the key characteristics of the data, some details are not accurately reproduced, as addressed in section 4.3. These are mainly attributed to the fact that this model was designed as simple as possible to focus on simulating the hypothesized origin of binaural sluggishness.

4.5 Conclusions

Psychoacoustic data sets that previously required very different binaural processing speeds can now be successfully simulated with the same model and without changing the temporal model parameters. Different effective processing speeds for the different tasks are facilitated by fast interaural cue encoding but a slow adaptation process to a change in the IPD statistics of the masker. Sluggishness kicks in when simulating detection in a masker with changing IPD statistics while for everything else the model is currently very fast. In our preceding study we demonstrated how a signal- and task-driven across-frequency interference of IPD statistics can resolve an apparent contradiction about the required filter bandwidth (Eurich *et al.*, 2022). The present study is an analogous strategy in the time domain to resolve an equally long-lasting apparent contradiction about the processing speed of the binaural system: It is very fast and very sluggish at the same time.

4.6 Acknowledgements

We thank Jörg Encke and Jonas Klug for fruitful discussions and comments on the manuscript. This work has received funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 352015383–SFB 1330 B4 and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 716800).

References

Akeroyd, M. A., and Bernstein, L. R. (**2001**). “The variation across time of sensitivity to interaural disparities: Behavioral measurements and quantitative analyses,” The Journal of the Acoustical Society of America **110**(5), 2516–2526, doi: `10.1121/1.1412442`.

Bernstein, L. R., and Trahiotis, C. (**2017**). “Binaural detection-based estimates of precision of coding of interaural temporal disparities across center frequency,” The Journal of the Acoustical Society of America **141**(5), 3973–3973, doi: `10.1121/1.4989060`.

Bernstein, L. R., Trahiotis, C., Akeroyd, M. A., and Hartung, K. (**2001**). “Sensitivity to brief changes of interaural time and interaural intensity,” The Journal of the Acoustical Society of America **109**(4), 1604–1615, doi: `10.1121/1.1354203`.

Bischof, N. F., Aublin, P. G., and Seeber, B. U. (**2023**). “Fast processing models effects of reflections on binaural unmasking,” Acta Acustica **7**, 11, doi: `10.1051/aacus/2023005`.

Boehnke, S. E., Hall, S. E., and Marquardt, T. (**2002**). “Detection of static and dynamic changes in interaural correlation,” The Journal of the Acoustical Society of America **112**(4), 1617–1626, doi: `10.1121/1.1504857`.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**). “Binaural processing model based on contralateral inhibition. I. Model structure,” The Journal of the Acoustical Society of America **110**(2), 1074–1088, doi: `10.1121/1.1383297`.

Buss, E., and Hall III, J. W. (**2011**). “Effects of non-simultaneous masking on the binaural masking level difference,” The Journal of the Acoustical Society of America **129**(2), 907–919, doi: `10.1121/1.3514528`.

Culling, J. F., and Mansell, E. R. (**2013**). “Speech intelligibility among modulated and spatially distributed noise sources,” The Journal of the Acoustical Society of America **133**(4), 2254–2261, doi: `10.1121/1.4794384`.

Culling, J. F., and Summerfield, Q. (**1998**). “Measurements of the binaural temporal window using a detection task,” The Journal of the Acoustical Society of America **103**(6), 3540–3553, doi: `10.1121/1.423061`.

Dietz, M., Ewert, S. D., Hohmann, V., and Kollmeier, B. (**2008**). "Coding of temporally fluctuating interaural timing disparities in a binaural processing model based on phase differences," Brain Research **1220**, 234–245, doi: `10.1016/j.brainres.2007.09.026`.

Encke, J., and Dietz, M. (**2022**). "A hemispheric two-channel code accounts for binaural unmasking in humans," Communications Biology **5**(1), 1122, doi: `10.1038/s42003-022-04098-x`.

Eurich, B., Encke, J., Ewert, S. D., and Dietz, M. (**2022**). "Lower interaural coherence in off-signal bands impairs binaural detection," The Journal of the Acoustical Society of America **151**(6), 3927–3936, doi: `10.1121/10.0011673`.

Gatehouse, S., and Akeroyd, M. (**2006**). "Two-eared listening in dynamic situations," International Journal of Audiology **45**(sup1), 120–124, doi: `10.1080/14992020600783103`.

Grantham, D. W. (**1982**). "Detectability of time-varying interaural correlation in narrow-band noise stimuli," The Journal of the Acoustical Society of America **72**(4), 1178–1184, doi: `10.1121/1.388326`.

Grantham, D. W., and Wightman, F. L. (**1979**). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," **65**, 1509–1517, doi: `10.1121/1.382915`.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics*, repr. ed ed. (Peninsula Publ, Los Altos Hills, Calif).

Hacker, M. J., and Ratcliff, R. (**1979**). "A revisted table of d′ for M-alternative forced choice," Perception & Psychophysics **26**(2), 168–170, doi: `10.3758/BF03208311`.

Hafter, E. R., and Dye, R. H. (**1983**). "Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number," The Journal of the Acoustical Society of America **73**(2), 644–651, doi: `10.1121/1.388956`.

Hafter, E. R., Dye, R. H., and Gilkey, R. H. (**1979**). "Lateralization of tonal signals which have neither onsets nor offsets," The Journal of the Acoustical Society of America **65**(2), 471–477, doi: `10.1121/1.382346`.

Hauth, C. F., and Brand, T. (**2018**). "Modeling Sluggishness in Binaural Unmasking of Speech for Maskers With Time-Varying Interaural Phase Differences," Trends in Hearing **22**, 233121651775354, doi: `10.1177/2331216517753547`.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a Gammatone filterbank," Acta Acustica united with Acustica **88**(3), 433–442.

Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). "Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments," The Journal of the Acoustical Society of America **104**(4), 2412–2425, doi: `10.1121/1.423773`.

Houtgast, T., and Plomp, R. (**1968**). "Lateralization Threshold of a Signal in Noise," The Journal of the Acoustical Society of America **44**(3), 807–812, doi: `10.1121/1.1911178`.

Just, D., and Bamler, R. (**1994**). “Phase statistics of interferograms with applications to synthetic aperture radar,” Applied Optics **33**(20), 4361, doi: `10.1364/AO.33.004361`.

Kolarik, A. J., and Culling, J. F. (**2009**). “Measurement of the binaural temporal window using a lateralisation task,” Hearing Research **248**(1-2), 60–68, doi: `10.1016/j.heares.2008.12.001`.

Kollmeier, B., and Gilkey, R. H. (**1990**). “Binaural forward and backward masking: Evidence for sluggishness in binaural detection,” The Journal of the Acoustical Society of America **87**(4), 1709–1719, doi: `10.1121/1.399419`.

Lüddemann, H., Riedel, H., and Kollmeier, B. (**2007**). “Logarithmic Scaling of Interaural Cross Correlation: A Model Based on Evidence from Psychophysics and EEG,” in *Hearing – From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey, Springer, Berlin, Heidelberg, pp. 379–388, doi: `10.1007/978-3-540-73009-5_41`.

McFadden, D. (**1966**). “Masking-Level Differences with Continuous and with Burst Masking Noise,” The Journal of the Acoustical Society of America **40**(6), 1414–1419, doi: `10.1121/1.1910241`.

McNemar, Q. (**1969**). *Psychological Statistics*, 4th ed ed. (Wiley, New York).

Reed, D. K., Dietz, M., Josupeit, A., and van de Par, S. (**2016**). “Lateralization of stimuli with alternating interaural time differences: The role of monaural envelope cues,” The Journal of the Acoustical Society of America **139**(1), 30–40, doi: `10.1121/1.4938018`.

Robinson, D. E., and Trahiotis, C. (**1972**). “Effects of signal duration and masker duration on detectability under diotic and dichotic listening conditions,” Perception & Psychophysics **12**(4), 333–334, doi: `10.3758/BF03207216`.

Ross, B., Miyazaki, T., Thompson, J., Jamali, S., and Fujioka, T. (**2014**). “Human cortical responses to slow and fast binaural beats reveal multiple mechanisms of binaural hearing,” Journal of Neurophysiology **112**(8), 1871–1884, doi: `10.1152/jn.00224.2014`.

Shinn-Cunningham, B., Best, V., and Lee, A. K. C. (**2017**). “Auditory Object Formation and Selection,” in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay, Springer Handbook of Auditory Research (Springer International Publishing, Cham), pp. 7–40, doi: `10.1007/978-3-319-51662-2_2;`.

Singh, R., and Bharadwaj, H. M. (**2021**). “Cortical Temporal Integration Window for Binaural Cues accounts for Sluggish Auditory Spatial Perception,” Preprint, doi: `10.1101/2021.12.14.472656`.

Siveke, I., Ewert, S. D., Grothe, B., and Wiegrebe, L. (**2008**). “Psychophysical and Physiological Evidence for Fast Binaural Processing,” Journal of Neuroscience **28**(9), 2043–2052, doi: `10.1523/JNEUROSCI.4488-07.2008`.

Stecker, G. C. (**2014**). “Temporal weighting functions for interaural time and level differences. IV. Effects of carrier frequency,” The Journal of the Acoustical Society of America **136**(6), 3221–3232, doi: `10.1121/1.4900827`.

Viemeister, N. F., and Wakefield, G. H. (**1989**). “Multiple looks and temporal integration,” The Journal of the Acoustical Society of America **86**(S1), S23, doi: `10.1121/1.2027422`.

Wilson, R. H., and Fugleberg, R. A. (**1987**). “Influence of Signal Duration on the Masking-Level Difference,” Journal of Speech, Language, and Hearing Research **30**(3), 330–334, doi: `10.1044/jshr.3003.330`.

Witton, C., Green, G. G. R., Rees, A., and Henning, G. B. (**2000**). “Monaural and binaural detection of sinusoidal phase modulation of a 500-Hz tone,” The Journal of the Acoustical Society of America **108**(4), 1826–1833, doi: `10.1121/1.1310195`.

Yabe, H., Winkler, I., Czigler, I., Koyama, S., Kakigi, R., Sutoh, T., Hiruma, T., and Kaneko, S. (**2001**). “Organizing sound sequences in the human brain: The interplay of auditory streaming and temporal integration11Published on the World Wide Web on 27 February 2001.,” Brain Research **897**(1), 222–227, doi: `10.1016/S0006-8993(01)02224-7`.

Yasin, I., and Henning, G. B. (**2012**). “The effects of noise-bandwidth, noise-fringe duration, and temporal signal location on the binaural masking-level difference,” The Journal of the Acoustical Society of America **132**(1), 327–338, doi: `10.1121/1.4718454`.

Yost, W. A. (**1985**). “Prior stimulation and the masking-level difference,” The Journal of the Acoustical Society of America **78**(3), 901–907, doi: `10.1121/1.392920`.

Zwicker, E., and Fastl, H. (**1999**). *Masking*, **22**, 61–110 (Springer Berlin Heidelberg, Berlin, Heidelberg), doi: `10.1007/978-3-662-09562-1_4`.

CHAPTER 5

A Computationally Efficient Model for Combined Assessment of Monaural and Binaural Audio Quality

This chapter is a formatted reprint of an identically titled manuscripted submitted to the *Journal of the Audio Engineering Society* on September 28th, 2023, with the citation style adapted to the author-year style used throughout the thesis.

The authors of the manuscript are:
Eurich, B.; Ewert, S. D.; Dietz, M.; Biberger, T.

Author contributions: BE, adapted the binaural model frontend first presented in chapter 3, re-implemented the monaural frontend presented in Biberger *et al.* 2018, developed the backend to assess combined monaural and binaural audio quality, computed the predictions, prepared the figures and wrote the manuscript. TB provided MATLAB code existing from previous publications to support the computation, analysis, and plotting of scores, organized the availability of the databases to be simulated, and participated in the design of the study and writing of the manuscript. MD and SE participated in the conceptual discussions underlying the model and in the improvement of the manuscript.

5.1 Abstract

Audio quality is an important aspect of hearing aids, hearables, and sound reproduction systems as the signal processing of such devices might alter the spectral composition or interaural differences of the original sound, and thus might degrade the perceived audio quality. Consequently, an audio quality model applicable to such devices requires to account for monaural and binaural aspects of audio quality. Flessner et al. (2019, IEEE/ACM Trans. Audio Speech Lang. Process. 2019, 27(7), 1112–1125) successfully predicted overall audio quality by combining a monaural and a binaural audio quality model, which is computationally expensive and thus limits the scope of application. In order to cover also time-critical applications, such as real-time control of algorithms in hearing technology, we present a computationally efficient audio quality model for overall quality predictions. The suggested model was evaluated with six databases including music and speech signals processed by loudspeakers and algorithms typically applied in modern hearing devices (e.g., acoustic transparency, feedback cancellation or binaural beamforming). The presented model achieved a high prediction performance, indicated by the mean Pearson correlation of 0.9 similar to the more complex model of Fleßner et al., while its calculation time is substantially lower by a factor of 70.

5.2 Introduction

Audio quality is an important aspect of many signal processing applications ranging from hearing devices to sound reproduction systems. For the evaluation of the perceived audio quality of algorithms or devices, listening tests are considered as the “gold standard”. These tests can be carried out as reference-free tests [e.g., ITU-T Rec. P. 800 1996], where listeners rate the audio quality of a processed speech or audio signal without any given unprocessed reference signal or as reference-based tests [e.g., (Munson and Gardner, 2005; Series, 2014)], comparing processed and unprocessed (reference) signals. Such listening test are typically time consuming, expensive and often require expert listeners to gain reliable quality judgements. To overcome these disadvantages, several instrumental audio quality measures have been developed (e.g., Moore and Tan (2004); Harlander *et al.* (2014); Biberger *et al.* (2018); Fleßner *et al.* (2019)). In addition to evaluating signal processing algorithms, instrumental quality measures can also be applied to control algorithms, provided they are computationally efficient.

One relevant field of application are wireless and smart headphones, in the following denoted as hearables. These devices have become increasingly popular because, in addition to their traditional use for listening to music and streaming audio, they offer signal processing features typically used in hearing aids to restore ambient sound for (hearing-impaired) listeners (Temme, 2019). The signal processing typically involved, such as noise suppression, beamforming, hear-through processing, nonlinear amplification, or attenuation, potentially alters the spectral composition or interaural differences of the original sound. This might be perceived by the listeners as spectral or spatial distortions, degrading the audio quality of signals. Hereby, hear-through processing aims at a natural (ideally acoustically transparent) representation of the external acoustical environment without perceivable distortions, similar to the sound impression with an open ear (without inserted device). This enables perceptually authentic conversations as well as awareness of the acoustic scene, both important in real life but also for augmented, mixed and virtual reality applications (Gupta *et al.*, 2020). Because the human auditory system is limited in its ability to resolve monaural (spectral and temporal) and binaural differences, such as interaural level and time differences (ILDs and ITDs), an authentic hear-through processing does not require the exact reproduction of the open-ear signal at the eardrum. A previous study (Biberger *et al.*, 2021) has shown that much of the distortion associated with the hear-through mode in hearables and smart headphones can be attributed to monaural, spectral coloration cues. However, degraded binaural cues, play a significant role in standard hearing device algorithms, such as binaural noise reduction and beamforming (Derleth *et al.*, 2021; Gößling *et al.*, 2021; Marquardt *et al.*, 2015; Doclo *et al.*, 2010). Given that binaural cues offer substantial advantages for speech intelligibility in realistic, complex acoustic conditions (Bronkhorst and Plomp, 1988, 1992; Bronkhorst, 2000; Hawley *et al.*, 2004), for sound localization (Blauert, 1996; Grothe *et al.*, 2010) as well as for listening effort (Rennies and Kidd, 2018), changes in (monaural) spectral coloration alone may not be a sufficient predictor for overall audio quality in such cases.

In the past, several monaural instrumental measures for the assessment of speech and audio quality have been developed (Harlander *et al.*, 2014; Huber and Kollmeier, 2006; Biberger *et al.*, 2018; Kates and Arehart, 2010, 2014, 2016; Beerends *et al.*, 2013; Rix *et al.*, 2001; Moore and Tan, 2004; Moore *et al.*, 2004), often designed for different specific applications, such as quality predictions for audio and speech codecs, hearing-aid signal processing, or loudspeaker and headphone distortions. In comparison to those monaural measures, only a few instrumental measures that capture binaural audio quality aspects have been developed (Flessner *et al.*, 2017; Schäfer *et al.*, 2013; Seo *et al.*, 2013; Takanen *et al.*, 2014; Manocha *et al.*, 2022), while such aspects are expected to be important in hearing devices (Marquardt *et al.*, 2015; Gößling *et al.*, 2021; Thiemann *et al.*, 2016; Yousefian *et al.*, 2014; Rohdenburg *et al.*, 2007; Doclo *et al.*, 2010; Derleth *et al.*, 2021) and (multi-channel) loudspeaker-based sound field reproduction (Toole, 1985; Gabrielsson and Lindström, 1985; Rumsey *et al.*, 2005). In a study by Rumsey and colleagues 2005, spatial fidelity accounted for approximately 30 % of the basic audio quality rating of degraded multichannel audio signals. Therefore they suggested to include spatial quality aspects in future perceptual models of sound quality. In the context of hearing aid processing, several recently suggested algorithms for noise reduction or de-reverberation were designed not only to improve speech intelligibility, but also to preserve binaural cues. Algorithms presented in Marquardt *et al.* (2015); Gößling *et al.* (2021) aimed at finding an optimal trade-off between noise reduction performance and the preservation of the interaural coherence for diffuse noise fields in order to maintain the spatial impression of the acoustical scene. The binaural de-reverberation algorithm presented in Jeub *et al.* (2010) was designed to suppress reverberation while maintaining binaural cues. Their listening test showed that for the objective assessment of such binaural-cue-preserving algorithms, instrumental quality measures require to account for spatial quality aspects, indicating whether the algorithm alters the spatial perception of the original sound.

One publicly available binaural instrumental audio quality measure is the binaural auditory model for audio quality [BAM-Q, Flessner *et al.* (2017)], an intrusive measure that is based on a perceptually motivated direction of arrival estimation model (Dietz *et al.*, 2011). It estimates spatial audio quality based on differences between the test and the reference signal in ILD, ITD, and an interaural coherence measure called interaural vector strength. In order to predict overall audio quality for signals impaired by monaural, binaural, or combined monaural and binaural distortions, Fleßner *et al.* (2019) suggested the instrumental audio quality measure MoBi-Q. Their model combines the outputs of the binaural BAM-Q and the monaural Generalized Power Spectrum Model for quality [GPSM$^{\mathrm{q}}$ Biberger *et al.* (2018)].

Their results were best described by the overall audio quality being determined by the lower quality aspect, i.e. either monaural or binaural quality. It represents an audio quality extension of the psychoacoustic and speech intelligibility model GPSM (Biberger and Ewert, 2016, 2017, 2022). The combined model has been shown to account for the distortions occuring in hearables (Biberger *et al.*, 2021).

So far, computational efficiency of binaural quality models was not specifically considered. For example, MoBi-Q combines the outputs of the independent GPSM^{q} and BAM-Q each including their own peripheral filtering stage, introducing computational redundancies. Therefore, potential for simplification and computation time reduction can be expected from unifying the monaural and binaural paths of the models. Particularly after it has been shown that peripheral filtering in the inner ear limits the bandwidth of both monaural and binaural processing bands in the same way (Mc Laughlin *et al.*, 2014; Dietz *et al.*, 2021; Eurich *et al.*, 2022).

Moreover, mammalian encoding of ITDs is best described as a two-hemisphere code (Grothe *et al.*, 2010; McAlpine *et al.*, 2001). Therefore, the complex correlation coefficient γ is a sufficient but compact formulation of the two-hemisphere code, reflecting coherence as the magnitude and IPD as the argument. It combines physiological plausibility with high predictive power in behavioral data and mathematical efficiency (Encke and Dietz, 2022; Eurich *et al.*, 2022).

Therefore, this study aims to provide a simplified and thus computationally more efficient version of MoBi-Q, consisting of the linear path of GPSM^{q} combined with γ and ILDs as binaural features, mimicking the perception of binaural cues. At the same time, it will be explored whether γ is suited to assess binaural audio quality.

5.3 Model Description

The architecture of the suggested quality measure allows to simultaneously analyze the binaural and monaural features in real time on a unified time scale, providing a frame-by-frame estimate of the binaural and monaural contributions to overall quality.

The relative contribution and perceptual range of the binaural features, γ and ILD, were first calibrated using a database of subjective quality ratings of the hear-through mode of hearables (Schepker *et al.*, 2020), following Biberger *et al.* (2021), since the model is aimed at applications in modern hearing and headphone technology. The model was then evaluated with six databases covering a broad range of monaural, binaural, and combined distortions. These databases include audio quality ratings on the acoustical transparency of binaural noise reduction algorithms, binaural magnification and adaptive feedback cancellation in hearing devices, loudspeaker distortions, and the acoustical transparency of hearing device prototypes. Finally, the performance of the proposed instrumental quality measure was compared to the more complex binaural quality measure BAM-Q and combined quality measure MoBi-Q.

Figure 5.1 shows the block diagram of the suggested efficient model for combined assessment of monaural and binaural audio quality (eMoBi-Q). The model requires processed (distorted) test and unprocessed reference signals with either one-channel (monaural) or two-channel audio signals (binaural) as input. In the following, the model frontend with joint preprocessing stages and the calculation of monaural and binaural features is explained, followed by the description of the backend where monaural and binaural features are combined to the final audio quality measure.

5.3.1 Front End

Preprocessing

Basilar membrane filtering of the left and right input signals was modeled by a linear fourth order gammatone filterbank (Patterson *et al.*, 1987; Holdsworth *et al.*, 1988), as implemented by Hohmann (2002). This results in 29 band-pass filtered signals with center frequencies between 315 Hz and 12.5 kHz that have equivalent rectangular bandwidths (ERB) according to (Glasberg and Moore, 1990). Processing stages in the monaural and binaural path process the bandpass signals in consecutive time frames of 400 ms. The time-frequency signal elements of the left and right ear signals are denoted as $l(n,p)$ and $r(n,p)$ for a time frame n and a frequency band p. A first-order lowpass filter with a 150 Hz cutoff frequency was applied to the envelope to model the limited sensitivity to envelope fluctuations (Kohlrausch *et al.*, 2000). The lowpass-filtered envelope affected the monaural spectral coloration feature, the ILD feature and the γ feature for frequency bands above 1300 Hz. However, no lowpass filtering was applied to the temporal fine-structure processing realised by the γ feature in frequency bands centred below 1300 Hz.

Monaural spectral coloration and loudness feature

The monaural feature was calculated by adopting the power spectrum path of the GPSM^q (Biberger *et al.*, 2018). In case of two-channel (binaural) input signals, left and right channels were concatenated. The Hilbert envelope, calculated for each of the complex-valued gammatone filterbank outputs, was filtered by a first-order lowpass filter with a 150 Hz cut-off frequency to account for the decrease of modulation sensitivity with increasing modulation frequency. The local DC power was extracted from the lowpass filtered envelope signals. It is half the squared mean of the envelope E across the time frame n of a frequency band p:

$$P(n,p) = \frac{\overline{E(n,p)}^2}{2} \tag{5.1}$$

Elements with a local DC power below the hearing threshold in quiet (iso, 2005) were set to that threshold.

As in (Biberger *et al.*, 2018), local power increments $\mathrm{SNR}_{\mathrm{incr}}$ were computed as[1]

$$\mathrm{SNR}(n,p)_{\mathrm{incr}} = \frac{P_{\mathrm{test}}(n,p) - P_{\mathrm{ref}}(n,p)}{P_{\mathrm{ref}}(n,p)} \tag{5.2}$$

and local power decrements $\mathrm{SNR}_{\mathrm{decr}}$ as

$$\mathrm{SNR}(n,p)_{\mathrm{decr}} = \frac{P_{\mathrm{ref}}(n,p) - P_{\mathrm{test}}(n,p)}{P_{\mathrm{test}}(n,p)}. \tag{5.3}$$

An upper limit of 13 dB was applied to each time-frequency element of $\mathrm{SNR}(n,p)_{\mathrm{incr}}$ and $\mathrm{SNR}(n,p)_{\mathrm{decr}}$, resulting in a dynamic range of 26 dB in total. Then $\mathrm{SNR}(n,p)_{\mathrm{incr}}$ and $\mathrm{SNR}(n,p)_{\mathrm{decr}}$ are averaged across time segments resulting in $\mathrm{SNR}(p)_{\mathrm{incr}}$ and $\mathrm{SNR}(p)_{\mathrm{decr}}$.

Binaural features

Two binaural features were extracted for each gammatone-filtered signal:

---

[1] Since the spectral coloration feature was adopted from GPSM^q, the term signal-to-noise ratio (SNR) was also used as historically established in, for example, the underlying GPSM, which predicts psychoacoustic masking and speech intelligibility (Biberger and Ewert, 2016). However, in the context of audio quality, “signal” and “noise” refer to “processed by device under test” and “unprocessed reference”, respectively. Thus, a high SNR means a high local power increment or decrement, which, although unintuitive, means strong distortion.

Complex correlation coefficient γ The complex-valued correlation coefficient was used because it conveniently combines information about both the interaural coherence $|\gamma|$, reflecting the perceptual compactness of a sound, and about the mean IPD as $\arg\{\gamma\}$, reflecting laterality. It is a mathematical formulation of the two-hemisphere channel code underlying neural encoding of interaural differences in mammals (Grothe *et al.*, 2010; McAlpine *et al.*, 2001), capturing temporal fluctuations in interaural phase. This feature and its assumption on filter bandwidth has been psychoacoustically validated by Encke and Dietz (2022), Eurich *et al.* (2022), Eurich and Dietz (2023) and Dietz *et al.* (2021).

The gammatone filterbank implementation (Hohmann, 2002) provides complex-valued outputs signals $l(n,p)$ and $r(n,p)$, utilized for computing the complex correlation coefficient γ:

$$\gamma(n,p) = \frac{\overline{l(n,p)^* r(n,p)}}{\sqrt{\overline{|l(n,p)|^2}\,\overline{|r(n,p)|^2}}} \tag{5.4}$$

where $\overline{\bullet}$ denotes the mean over the duration of the time frame. For frequency bands with center frequencies below 1300 Hz, γ operates on the temporal fine structure of the bandpass signals, while above of 1300 Hz it operates on their Hilbert envelopes. This mimics the sensitivity to IPDs in the temporal fine structure at low frequencies as encoded by the human medial superior olive (MSO) in combination with the sensitivity to IPDs in the envelope at higher frequencies as encoded by the lateral superior olive (LSO) (Remme *et al.*, 2014; Klug and Dietz, 2022). As in Eurich *et al.* (2022); Eurich and Dietz (2023), Fisher's z transform was applied to the coherence (i.e. $|\gamma|$) to normalize the variance and to account for the increasing sensitivity to changes in coherence towards unity. To avoid infinite sensitivity, γ was multiplied by 0.9 (Eurich *et al.*, 2022; Eurich and Dietz, 2023).

Interaural level differences Interaural level differences (ILDs) were extracted as the logarithmic power ratio between left and right signals:

$$\mathrm{ILD}(n,p) = 10\log\left(\frac{P_l(n,p)}{P_r(n,p)}\right) \tag{5.5}$$

The model's sensitivity to binaural distortions was obtained as the difference between the frontend outputs of reference and test signals, denoted as d':

$$d'_\gamma(n,p) = |\gamma_{\mathrm{ref}}(n,p) - \gamma_{\mathrm{test}}(n,p)| \tag{5.6}$$

$$d'_{\mathrm{ILD}}(n,p) = |\mathrm{ILD}_{\mathrm{ref}}(n,p) - \mathrm{ILD}_{\mathrm{test}}(n,p)| \tag{5.7}$$

An upper limit of 10 dB [calibrated to the hear-through-mode database (Schepker *et al.*, 2020)] was applied to $d'_{\mathrm{ILD}}(n,p)$ [cf. BAM-Q, Flessner *et al.* (2017)], to mimick the perceptual saturation of laterality and to avoid disproportionately large ILDs at moments of very low one-sided DC power.

5.3.2 Backend

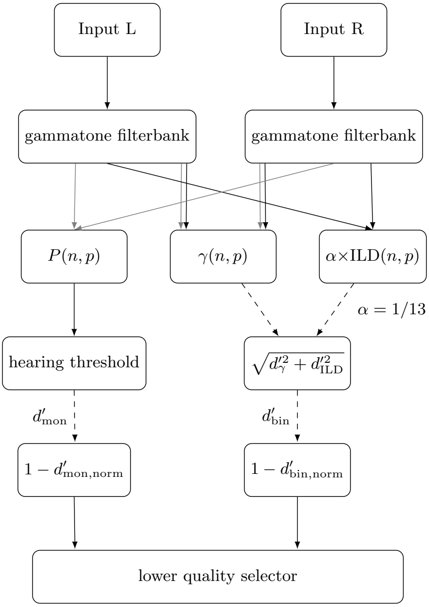

Figure 5.1: Block diagram of the proposed model. In both the monaural (local DC power $P(n,p)$ and binaural ($\gamma(p,n)$, $\mathrm{ILD}(p,n)$) paths, the frequency channels p are combined in an optimal manner. The n consecutive 400 ms time frames are combined in an optimal manner for the binaural features and averaged for the spectral coloration feature. Gray lines denote envelope-lowpass filtering of the audio signals, dashed lines denote that the discriminability d' was obtained from comparing a test signal with a reference signal.

For $d'_\gamma(n,p)$ and $d'_{\mathrm{ILD}}(n,p)$, information was optimally combined across time frames n and frequency bands p, i.e. assuming a linear, independent combination:

$$d' = \sqrt{\sum_n \sum_p d'(n,p)^2}. \tag{5.8}$$

The weighted optimal combination of the two binaural features' sensitivity indices gives the output of the binaural model path:

$$d'_{\mathrm{bin}} = \sqrt{d'^2_\gamma + \alpha\, d'^2_{\mathrm{ILD}}} \tag{5.9}$$

where the relative weight $\alpha = 1/13$ of the ILD feature was calibrated using the database on the hear-through mode of hearables (Schepker *et al.*, 2020).

Adopted from Biberger *et al.* (2018), the monaural increment and decrement SNRs, $\mathrm{SNR}(p)_{\mathrm{incr}}$ and $\mathrm{SNR}(p)_{\mathrm{decr}}$, were combined by taking the mean for each auditory filter, resulting in $\mathrm{SNR}(p)_{\mathrm{mon}}$. These monaural SNRs were then optimally combined across frequency bands providing the single-valued $\mathrm{SNR}_{\mathrm{mon}}$ to which a logarithmic transformation was applied with lower and upper bounds as resulting from the 26 dB dynamic range (Biberger *et al.*, 2018):

$$d'_{\mathrm{mon,\ lim}} = \min(\max(10\log(\mathrm{SNR}_{\mathrm{mon}}) + 10, 0), 26). \tag{5.10}$$

The dynamic range of the binaural d'' is limited by a lower bound of zero and an upper bound of 23, calibrated to the hear-through-mode database (Schepker *et al.*, 2020). The perceptual range of both model paths was normalized to $d'_{\mathrm{norm}} \in [0;\ 1]$: While the sensitivity indices of the model, d', represent the perceptual distance between reference and test signals, the predicted audio quality was obtained as $1 - d'_{\mathrm{norm}}$.[2] This allows to adopt the linear monaural frontend from Biberger *et al.* (2018) and combine it with the new binaural frontend without further calibration.

In psychoacoustic detection tasks, monaural and binaural cues are usually best described by an optimal combination (Encke and Dietz, 2022). However, Fleßner

---

[2] The Weber law suggests that a logarithmic d' axis is more likely to reflect perception than a linear axis (Heil and Friedrich, 2023). This would suggest associating $log(1/d')$ with audio quality rather than $1 - d'$. However, a backend based on $log(1/d')$ did not give better results and requires a modification of the $d_{\mathrm{mon,lim}}$ adopted from (Biberger *et al.*, 2018). Therefore, in this work, a linear association of d' with audio quality is used, which provides simplicity paired with performance. For future backends involving, e.g., a neural network, a logarithmic association of d' and audio quality may be preferred.

et al. (2019) concluded from the combination functions tested for MoBi-Q that the overall audio quality is dominated by the lower quality aspect. For eMoBi-Q, selecting the lower quality component, i.e. monaural or binaural, also yielded the better results than an optimal combination (not shown). Therefore, in order to provide a simple but well performing combination, the lower quality component was selected as the overall quality rating. However, the features provided in this model can also be used with other backends (see section 5.6.3). This combined version of the model was used to predict the subjective ratings of the seven databases described below. Additionally, the performance of the monaural and binaural paths in isolation was compared to previous models, which is discussed in section 5.6.1.

5.4 Databases

The hear-through mode database database of Schepker *et al.* (2020) was used for calibration of the relative weight of the binaural features, i.e. γ and ILDs, as well upper bound of ILD cues and the upper bound of the binaural path, cf. Biberger *et al.* (2021). Six further databases covering a broad variety of monaural, binaural and combined monaural and binaural distortions as they typically occur in loudspeakers and hearing technology were used to evaluate the “calibrated" model.

The hear-through mode database, used for calibration, was taken from the study of Schepker et al. 2020 (Schepker *et al.*, 2020). The database consists of 120 speech (female, male) and music (jazz, piano) items, sampled at 48 kHz. The study aimed to assess the audio quality of various hearables, including six commercial devices and three research devices, in the hear-through mode. To achieve this, recordings were made using a dummy head equipped with the hearables in a laboratory environment with moderate room reverberation (T60 $\approx$ 0.45 s) to assess the devices in realistic but controlled acoustic conditions. Four audio signals were recorded for three playback directions (azimuths of 0°, 90°, 225°) with loudspeakers placed at a distance of approximately 2 m from the dummy head and adjusted in height to be at ear level with the dummy head. The dummy head’s open-ear recordings served as the reference signals, ensuring that the sound transmission to the eardrum through the hearable devices matched the acoustic transparency of the open ear reference. The occluded ear was used as anchor signal. The subjective evaluation of the hearables was conducted with 17 NH participants by employing a MUSHRA-like framework.

The following six databases were used for model evaluation. The subjective quality ratings for all these databases were measured in headphone experiments with participants who had normal hearing (NH) in sound-isolated booths.

5.4.1 Binaural Distortions

The database by Flessner *et al.* (2017) has 114 items, consisting of speech, music, and pink noise signals with a duration of 10 s. The reference signals were diotic and thus perceived in the middle of the head as a narrow spatial image. The test signals were manipulated in ILDs and ITDs to change the perceived apparent source width, listening envelopment and the direction of arrival of the sound source. The listeners rated the perceived difference between a reference and various test signals on a numerical rating scale ranging from 100 (“no difference”) to 0 (“very strong difference”) by using a procedure similar to the MUSHRA (Multiple Stimulus with Hidden Reference and Anchor) method.

The binaural magnification database, including 8 items, sampled at 44.1 kHz, was taken from Flessner *et al.* (2017) and comprises binaural hearing aid algorithms (Kollmeier and Peissig, 1990), that magnifies binaural ILD- and ITD-cues to improve the spatial separation between sound sources. The algorithm was applied to one speaker in a conversation scenario who talks with another (unprocessed) speaker. Such processing shifts the perceived location of the processed speaker, while the spatial position of the other talker does not change. In the unprocessed reference signal both speakers were perceived in front of the receiver. Different degrees of magnifications were tested and 10 NH listeners rated the overall difference between the reference and the test signals by using a procedure similar to MUSHRA.

The database of Gößling *et al.* (2021) contains 32 speech items, sampled at 16 kHz, and a duration of about 7 s. In their study, Gößling et al. measured the performance of six noise reduction algorithms based on the binaural minimum-variance-distortionless-response (MVDR) beamformer, that compromise between noise reduction performance and preservation of the interaural coherence for diffuse noise fields. A MVDR beamformer with optimal processing strategy (MVDR-OPT) that reduces the signal-to-noise ratio between the speech and noise component but perfectly preserves the interaural coherence of the diffuse noise component was used as the reference signal. The anchor signal was obtained by averaging the left and the right output signals of the MVDR-OPT algorithm, resulting in a monaural signal. Consequences of such algorithms on the perceived audio quality were assessed for anechoic and echoic (cafeteria) room conditions. Eleven NH listeners rated the perceived audio quality between the test and the reference signals by using a MUSHRA-like procedure.

5.4.2 Monaural Distortions

The loudspeaker database, taken from Biberger *et al.* (2018), consists of 336 items (sampled at 44.1 kHz), based on the ratings of 10 well-trained NH listeners ("expert listeners") for the perceived overall sound quality difference between a high-quality three-way reference loudspeaker and 59 low-to-mid quality three-way and two-way test speaker systems playing 15 music excerpts (20-30 s). All loudspeakers were digitally equalized in order to evaluate quality differences between test loudspeakers with digitally compensated frequency response and a high-quality three-way reference loudspeaker. The played-back music signals were recorded by a dummy head (Neutric Cortex MK2). The perceived sound quality differences between reference and test signals were rated by using a quasi-continuous rating scale ranging from 0 (imperceptible differences) to 4 (significant differences).

The adaptive feedback cancelation (AFC) database was taken from the study of Nordholm *et al.* (2018). It consists of 60 diotic items, based on speech and music material, sampled at 16 kHz. All signals were recorded using a microphone placed in the right ear of a dummy head in an anechoic chamber for two different sound source positions (azimuths of 0° and 90°), resulting in four audio signals (2x speech and 2x music). Nordholm et al. examined four AFC algorithms using four signals and three signal segments (initial and re-convergence phase, steady-state phase). Signals processed with an ideal feedback cancelation algorithm (with perfect a-priori knowledge about the feedback path) served as reference signals, while signals processed without feedback cancellation served as anchor signals. Subjective quality ratings from 15 NH subjects were obtained using the Multiple Stimulus with Hidden Reference and Anchor method (MUSHRA; ITU-R BS.1534-1, 2003).

5.4.3 Combined Distortions

The acoustic transparency database, taken from the study of Schepker *et al.* (2019), encompasses 140 speech and music items, sampled at 48 kHz. The study aimed to evaluate the audio quality of a real-time hearing device prototype designed for achieving acoustically transparent sound reproduction by applying feedback suppression using a null-steering beamformer and individualized equalization of the sound pressure at the eardrum. The evaluation was conducted under various recording room conditions, including three different reverberation times (T60 ≈ 0.35 s, 0.45 s, 1.4 s) and three incoming signal directions (azimuths of 0°, 90°, 225°). For the recording process, a dummy head equipped with the hearing devices was utilized. The open-ear recordings from the dummy head served as the reference signals to establish acoustical transparency. A total of 15 NH listeners were involved in the study, and they employed a MUSHRA-like procedure to rate the perceived overall sound quality of each stimulus relative to the reference signal (open-ear).

5.5 Results

Distortion Type	Study	Database	r_{Pearson}	r_{rank}
binaural	Fleßner et al. 2017	Artificial distortions	0.85 (0.85)	0.85 (0.85)
	Fleßner et al. 2017	Magnification hearing aid algorithm	0.96 (0.96)	0.95 (0.95)
	Gößling et al. 2020	MVDR-based algorithms	0.89 (0.98)	0.80 (0.95)
monaural	Biberger et al. 2018	Loudspeakers	0.86 (0.91)	0.90 (0.87)
	Nordholm et al. 2018	Adaptive feedback cancelation	0.99 (0.99)	0.98 (0.98)
combined	Schepker et al. 2019	Acoustically transparent hearing device	0.86	0.85
	Schepker et al. 2020	Calibration: Hear-through mode	0.90	0.90

Table 5.1: Results: Performance in terms of Pearson linear correlation coefficients r_{Pearson} and Spearman rank correlation coefficients r_{rank} between subjective and objective ratings for the seven databases predicted by the proposed combined monaural and binaural model eMoBi-Q. Results obtained with the binaural model path in isolation (for the databases on binaural distortions) or monaural model path in isolation (for the databases on monaural distortions) are given in parantheses.

Prediction performance is characterized by two performance measures: Accuracy is quantified by the Pearson linear correlation coefficient r_{Pearson}, monotonicity by the Spearman rank coefficient r_{rank}. Results are summarized in Table 5.1. There, r_{Pearson} and r_{rank} for each database are given as predicted by the combined monaural and binaural quality model (eMoBi-Q). Additionally, scores in parantheses denote the performance obtained by the binaural features in isolation (for binaural distortion databases) or the spectral coloration feature in isolation (for monaural distortion databases).

Figure 5.2 shows subjective quality scores and objective scores for eMoBi-Q for the

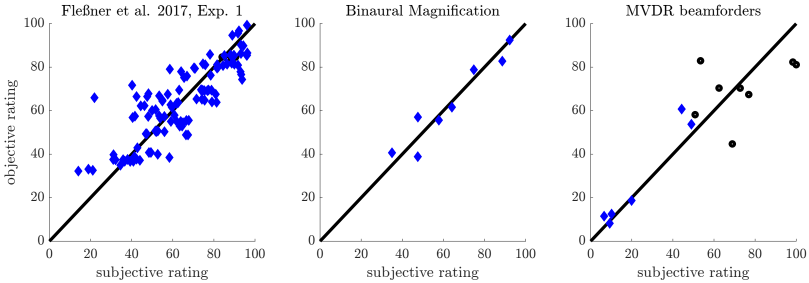

Figure 5.2: Subjective and objective quality scores for the databases on binaural distortions. Black circles denote conditions determined by the monaural path, i.e., spectral coloration based measure, being lower than the binaural distortion measure, blue diamonds denote those determined by a lower binaural measure. *Left panel:* Database by Fleßner et al. 2017 on artificial distortions in ITDs, ILDs and HRTFs; *middle panel:* Binaural magnification database by Fleßner et al. 2017; *right panel:* database on noise reduction algorithms based on the binaural minimum-variance-distortionless-response (MVDR) beamformer by(Gößling *et al.*, 2021).

binaural distortions in three databases (Flessner *et al.*, 2017; Gößling *et al.*, 2021). In the Figs. 5.2 - 5.4, subjective and objective, i.e., instrumentally assessed quality scores are given on the abscissa and on the ordinate, respectively. Black circles and blue diamonds denote predictions determined by lower spectral and binaural features, respectively. Table 5.1 lists the r_{Pearson} and r_{rank} coefficients between subjective and objective scores as obtained for eMobi-Q.

For the calibration database eMoBi-Q achieved $r_{\text{Pearson}} = 0.9$ and $r_{\text{rank}} = 0.9$. eMoBi-Q performed well the for the artificial binaural distortions in the database by Fleßner et al. ($r_{\text{Pearson}} = 0.85$, $r_{\text{rank}} = 0.85$) and gave accurate predictions for the magnification hearing aid algorithm, indicated by $r_{\text{Pearson}} = 0.96$ and r_{rank} 0.95, as well as for the MVDR-based algorithms ($r_{\text{Pearson}} = 0.89$, $r_{\text{rank}} = 0.8$). Table 5.1 shows that for these databases prediction performance increases when only the binaural path of eMoBi-Q is used.

In Figure 5.3 eMoBi-Q scores are plotted over subjective quality scores for the monaural distortions in the loudspeaker and adaptive feedback cancelation databases. For both databases eMoBi-Q provided good quality predictions for the loudspeaker database ($r_{\text{Pearson}} = 0.86$, $r_{\text{rank}} = 0.88$) and very accurate predictions for the adaptive feedback cancelation database ($r_{\text{Pearson}} = 0.99$, $r_{\text{rank}} = 0.98$) for the loudspeaker

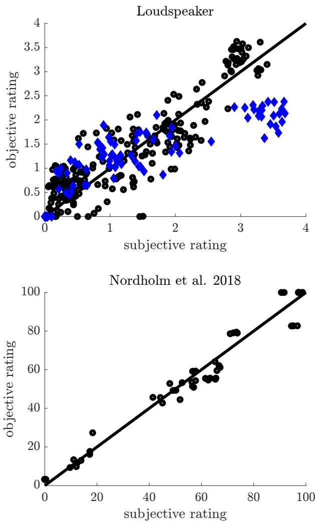

Figure 5.3: Subjective and objective quality assessments for the databases on monaural distortions. *Top panel:* Loudspeaker database; *Bottom panel:* adaptive feedback cancellation (AFC) database

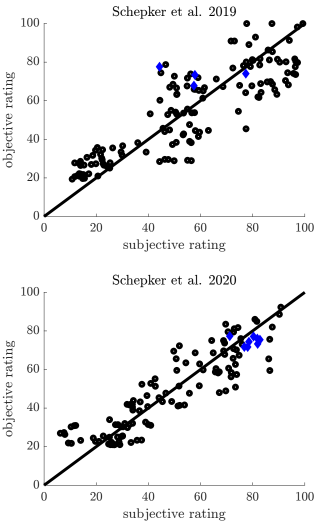

Figure 5.4: Subjective and objective quality assessments for the databases on combined monaural and binaural distortions. *Top panel:* Acoustic transparency database; *Bottom panel:* database on the quality of the hear-through mode of hearables, used for calibration of the binaural path.

and adaptive feedback cancellation databases, respectively. The prediction performance of eMoBi-Q for combined monaural and binaural distortions are shown in Figure 5.4. Next to the hear-through-mode database used for calibration of the binaural path, eMoBi-Q also replicated the ratings on the acoustic transparency of hearing aid prototypes (Schepker *et al.*, 2019) very well ($r_{\text{Pearson}} = 0.86$, $r_{\text{rank}} = 0.85$).

Without further optimization and parameter adjustment procedures, the presented combined model eMoBi-Q achieved average r_{Pearson} and r_{rank} coefficients between subjective ratings and objective model ratings of 0.9 and 0.89, respectively, for seven databases.

5.6 Discussion

The presented efficient model for combined assessment of monaural and binaural audio quality (eMoBi-Q) was shown to predict a range of monaural, binaural and combined distortions well. The involved features are the complex correlation coefficient γ which incorporates interaural coherence ($|\gamma|$) characterizing compactness and the IPD ($\arg\{\gamma\}$), ILD representing laterality, and DC power. With the model structure being transparent and simple, developers can incorporate the features into their analyses according to their own requirements.

5.6.1 Comparison to other instrumental quality measures

An instrumental quality measure intended to predict overall audio quality requires to capture aspects that degrade monaural and binaural audio quality. To assess the power of the auditory cues analyzed in eMoBi-Q, the prediction performance of the isolated monaural and binaural path of eMoBi-Q are in the following compared to existing monaural and binaural instrumental quality measures. Besides an adequate representation of monaural and binaural cues, the combination of such cues is also important to gain reasonable overall quality outcomes. Therefore, eMoBi-Q is additionally compared to an existing instrumental measure for overall audio quality.

Binaural measures

One goal in this study was to assess whether the simplistic and computationally efficient binaural auditory model of Eurich *et al.* (2022) is suitable to predict binaural audio quality. For that reason, the prediction performance of the binaural path of eMoBi-Q was compared to that of the established binaural audio quality model BAM-Q (Flessner *et al.*, 2017) for the three binaural databases in this study. As shown in Fig. 5.5 for the databases for binaural magnification and MVDR beamformers, the binaural path of eMoBi-Q has a prediction performance comparable to BAM-Q, which is also indicated by similar r_{Pearson} and r_{rank} above 0.9 for both models. However, for the database of Flessner *et al.* (2017), the prediction performance of the binaural path of eMoBi-Q ($r_{\text{Pearson}} = 0.85$, $r_{\text{rank}} = 0.85$) is lower than that of BAM-Q ($r_{\text{Pearson}} = 0.93$, $r_{\text{rank}} = 0.93$). Given that BAM-Q has been trained on the Fleßner database, it is not surprising that BAM-Q outperforms eMoBi-Q for

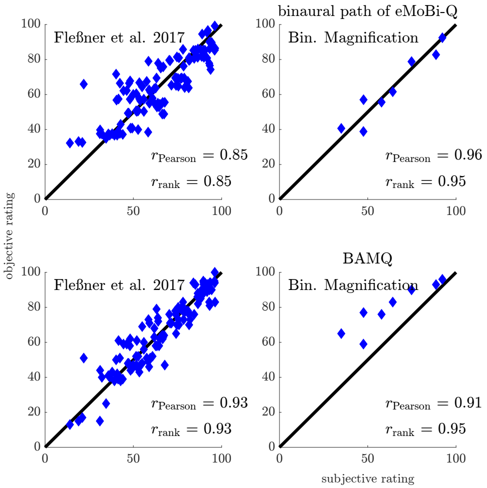

Figure 5.5: Performance comparison of the binaural path of the present eMoBi-Q in isolation (left column) with the established binaural quality model BAM-Q (right column). The used databases on binaural distortions are the data from experiment 1 in Fleßner et al. 2017, the binaural magnification database and the database of binaural cue preservation in B-MVDR beamformers from Gößling et al. 2020.

that database. The features extracted by BAM-Q – ITD, ILD and interaural vector strength (IVS) – are related to the features of eMoBi-Q, γ and ILD. The backend of BAM-Q, however, involves the "multivariate adaptive regression splines" [MARS, Friedman (1991); Jekabsons (2011)] consisting of forward and backward passes to fit the relative importance of the three features to the data, as well as further computations to obtain the quality ratings. In the present binaural model, however, $1 - d'_{\mathrm{norm}}$ is directly used as binaural quality rating. The proposed model can serve as a basis for potentially more elaborate backends to further optimize prediction accuracy. However, when the relative contribution of the ILD feature is increased to $\alpha = 1/8$, performance of the binaural path of eMoBi-Q for the Fleßner database becomes closer to BAM-Q ($r_{\mathrm{Pearson}} = 0.88$, $r_{\mathrm{rank}} = 0.89$). In a nutshell, the similar overall performance of the two models highlights the strength of the simplistic binaural path of eMoBi-Q and the suitability of the complex correlation coefficient γ for binaural quality assessment. Therefore, the binaural path of eMoBi-Q could also provide a useful binaural extension for other monaural audio quality models.

Monaural measures

The subjective quality ratings in the databases on loudspeakers and adaptive feedback cancellation were well replicated by the current eMoBi-Q model as indicated by r_{Pearson} values of 0.86 and 0.98, respectively. eMoBi-Q and the isolated monaural path of eMoBi-Q achieved the same prediction performance for the database on adaptive feedback cancelation (compare results without and with parenthesis in Table 5.1), while for the (dichotic) loudspeaker database, eMoBi-Q performed slightly worse then the isolated monaural path of eMoBi-Q. This is due to the cue redundancy in the monaural and binaural features (see below). Specifically, interaural coherence cues (i.e. $|\gamma|$) are present as loudspeaker database compares recordings in rooms. The studies of Biberger et al. 2018; 2021 demonstrated that for the loudspeaker and adaptive feedback cancellation databases, accurate predictions of the perceptual effects of spectral distortions are important. Therefore, the naturalness model (Moore and Tan, 2004), HASQIv2 (Kates and Arehart, 2014), and GPSM$^{\text{q}}$ (Biberger *et al.*, 2018) each explicitly accounting for spectral differences between reference and test signals, were used as monaural comparison models. For the loudspeaker database and the adaptive feedback cancelation database, eMoBi-Q performs similar to the naturalness model, HASQIv2 and GPSM$^{\text{q}}$ (loudspeaker database: r_{Pearson} values of 0.85, 0.8, and 0.9; adaptive feedback cancelation database: $r_{\text{Pearson}} = 0.95$ for all three comparison models).

Combined measures

Evaluating the binaural and the monaural path of eMoBi-Q in isolation has shown that binaural and monaural cues in hearing devices and loudspeakers are generally well predicted. This gives developers the choice of using the paths in isolation or in combination.

Biberger *et al.* (2021) tested the combination of GPSM$^{\text{q}}$ and BAM-Q (Flessner *et al.*, 2017) with the acoustic transparency database (Schepker *et al.*, 2019) (see section 5.4.3). While GPSM$^{\text{q}}$ alone performed well ($r_{\text{Pearson}} = 0.87$; $r_{\text{rank}} = 0.86$), performance was slightly reduced when it was combined with BAM-Q in MoBi-Q ($r_{\text{Pearson}} = 0.83$; $r_{\text{rank}} = 0.80$). This is not the case for eMoBi-Q, which performed equally well as GPSM$^{\text{q}}$ ($r_{\text{Pearson}} = 0.88$; $r_{\text{rank}} = 0.87$)

An even more significant detrimental impact of BAM-Q combined with GPSM$^{\text{q}}$

was observed for the hear-through mode database (Schepker *et al.*, 2020) (MoBi-Q: $r_{\mathrm{Pearson}} = 0.79$; $r_{\mathrm{rank}} = 0.81$; GPSM$^{\mathrm{q}}$: $r_{\mathrm{Pearson}} = 0.92$; $r_{\mathrm{rank}} = 0.91$). Given that eMoBi-Q was calibrated on the hear-through mode database (Schepker *et al.*, 2020) it seems plausible that it achieved a better performance ($r_{\mathrm{Pearson}} = 0.90$; $r_{\mathrm{rank}} = 0.90$) than MoBi-Q without any a-priori knowledge about that database.

The reduced prediction performance of the combined model compared to the monaural or binaural path in isolation can be explained by binaural distortions also being reflected in spectral distortions. Because ILDs are extracted as the logarithmic power ratio between the left and right bandpass signals, interaural differences in DC power are detected by both the ILD and the DC-power feature of eMoBi-Q. Furthermore, as discussed by Fleßner *et al.* (2019) and Biberger *et al.* (2021), the way of combining monaural and binaural paths has a major impact on the overall predicted quality and carries the risk of obtaining a large number of degrees of freedom, overfitting, and significant degradation of prediction performance. For this reason and for the sake of simplicity, for eMoBi-Q, no specific weighting of the monaural or binaural paths was used.

As one result of Fleßner *et al.* (2019) was that the lower quality component determines the overall quality, this was applied to eMoBi-Q. The result is a lean combined monaural and binaural instrumental quality measure with less degrees of freedom and which at the same time achieves a slightly higher performance than the more complex MoBi-Q on combined distortions considered in this study.

5.6.2 Computational efficiency

The goal was to provide a computationally efficient audio quality assessment model that can serve as both a real-time hearing device control and a development tool. For a 1 s two-channel audio signal, the model's signal processing takes about 257 ms For comparison: The current (unoptimized) implementation MoBi-Q (Fleßner *et al.*, 2019), needs 17 s.[3] One obvious redundancy in MoBi-Q are two separate peripheral filter stages for the binaural and monaural model, and thus eMoBi-Q uses a single, common filterbank. Because of the low computational complexity of the monaural and binaural feature calculation in eMoBi-Q, the peripheral filterbank requires 48 % of the run time, and the subsequent envelope lowpass filtering for a further 23 %,

---

[3] The models were run in MATLAB on an Intel(R) Core(TM) i7-8565U CPU @ 1.80 GHz machine, using a single thread.

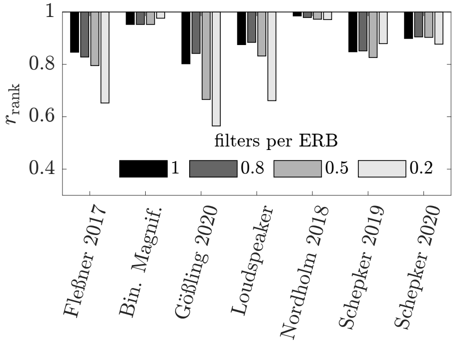

Figure 5.6: Performance of the presented eMoBi-Q in terms of prediction monotonicity (r_{rank}) for the seven considered databases for different spacings of the frequency bands. While the lower- and uppermost center frequencies are kept constant, the distance between center frequencies is increased. A lower number of frequency bands reduces computational load. Results given in Figs. 2 to 5 use 1 filter per ERB.

meaning that approximately 71 % of the total run time is spent in this initial stage. Reducing the number of frequency bands can therefore further reduce the computational load. To explore this potential, the model was evaluated with a reduced number of frequency bands. The lowest and highest center frequencies were kept constant at 315 Hz and 12500 Hz respectively, while the density of the frequency bands in between was reduced from 1 filter per ERB (default) to 0.8, 0.5 and, as an extreme case, 0.2 filters per ERB. With the filter bandwidth unadjusted, this led to a reduction in filter overlap and, in extreme cases, to the neglect of frequency ranges between the filters. With 0.5 filters per ERB, the run time was reduced from 257 ms to 121 ms which means the run time is approximately proportional to the number of frequency bands. The resulting performance in terms of their r_{rank} between subjective data and model predictions is shown in Fig. 5.6. Depending on the individual database, low to moderate performance losses were observed for 0.8 and 0.5 filters per ERB. Only for the database Gößling *et al.* (2021), a more significant loss was observed at 0.5 filters per ERB. For the extreme case of 0.2 filters per ERB, however, substantial performance losses were observed for three of the seven databases (binaural calibration, binaural magnification and loudspeaker database). We hypothesize that, based on the used set of seven databases, distortions that occur in one frequency band are likely to also occur in at least one neighboring frequency band. Thus, even if the sensitivity of the model is not constant over the entire frequency range (it is constant for the standard density of one filter per ERB, where transfer functions cross at their 3-dB-down points), a large part of the distortions that determine the subjective ratings are captured. Compensating the lower density of frequency bands with larger filter bandwidths led to more substantial performance losses (not shown). The more significant loss for the database by Gößling *et al.* on binaural cue preservation in MVDR beamformers, however, shows that binaural audio quality in such applications relies on cues that are not necessarily represented in adjacent frequency regions. We conclude that for some time-critical applications, such as real-time evaluation, it may be useful to use the model with a reduced number of frequency bands. However, in order to maintain generalizability to different stimuli with different bandwidths, it is recommended that the center frequencies of the remaining filters cover a wide range, such as 315…12500 Hz.

5.6.3 Limitations and reasonable model extensions

Besides the shown range of distortion types that are well captured by the presented model, there are also distortion types the model is not expected to be accounted for: The presented model does not include a feature to capture nonlinear distortions, which makes the model unsuitable to evaluate the audio quality of, e.g., audio codecs. Furthermore, distortions such as spectral subtraction, introducing musical tones, are not exptected to be accurately detected by the current version of the model without modulation filters.

The frame length of 400 ms was chosen as the focus was on detecting realistic binaural distortions (Flessner *et al.*, 2017; Fleßner *et al.*, 2019; Biberger *et al.*, 2021) and computational efficiency. Fast dynamic binaural distortions, such as phasewarp (i.e. a binaural beat created by an interaural spectrum shift) are, however, not detected. A future version could possibly include fine-structure-based feature extraction as used in Eurich and Dietz (2023), which would increase sensitivity to such mostly artificial distortions at the cost of higher computational load.

To address the redundancy of monaural and binaural cues, which partially results in performance degradation when the monaural DC power path is added to the binaural path and vice versa, a unified monaural and binaural path could be developed for a future version. Alternatively, ILDs and ITDs could be canceled out in the DC power path, as in MoBi-Q. Also, a sophisticated procedure to fit the relative weighting of the model features could potentially slightly improve performance. In the study of Qiao *et al.* (2022), a simple neural network was trained to map the monaural and binaural features of MoBi-Q for timbral, spatial, and overall quality. For their test databases, containing signals processed by binaural rendering algorithms and ambisonics reproduction, such mapping provided more accurate predictions than the original feature combination suggested in MoBi-Q. Thus, replacing the straightforward combination of monaural and binaural features in eMoBi-Q by a carefully trained neural network might also improve the prediction performance. However, the focus of this model was on efficiency, simplicity, and, considering the few degrees of freedom, generalizability.

5.7 Conclusion

A computationally efficient and lean instrumental measure for combined monaural and binaural audio quality assessment was presented. While a number of monaural instrumental quality measures has been established in the past, tools for assessing binaural aspects of audio quality are limited, although spatial cue preservation is important for, e.g., binaural hearing aids and sound-field reproduction. The presented model is a simplified version of MoBi-Q, providing a lean structure and a new, compact binaural path. The predictive power of the presented model was shown to be comparable with more computationally complex quality models for seven databases involving a range of monaural, binaural and combined distortions. Due to the simple structure, the resulting computational efficiency, and the unified analysis timescales of the monaural and binaural paths, the model is suitable for a range of applications. It has the potential for real-time control of algorithms in, e.g., hearing aids, but can also be used as an analysis tool for developers to monitor perceptually relevant distortions. The model will be publicly available from the University of Oldenburg and will be part of the Auditory Modeling Toolbox (Majdak *et al.*, 2022).

5.8 Acknowledgement

Thomas Biberger and Stephan Ewert were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — Project-ID 352015383–SFB 1330 A2. Bernhard Eurich and Mathias Dietz were funded by Deutsche Forschungsgemeinschaft — Project-ID 352015383–SFB 1330 B4 and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 716800).

References

(**2005**). “ISO (2005). 389-7, Acoustics-Reference Zero for the Calibration of Audiometric Equipment. Part 7: Reference Threshold of Hearing Under Free-Field and Diffuse-Field Listening Conditions (International Organization for Standardization, Geneva, Switzerland),” .

, ITUT, R. (**1996**). “ITU-T Rec. P. 800: Methods for Subjective Determination of Transmission Quality. Int. Telecomm. Un., Geneva,” .

Beerends, J. G., Schmidmer, C., Berger, J., Obermann, M., Ullmann, R., Pomy, J., and Keyhl, M. (**2013**). “Perceptual Objective Listening Quality Assessment (POLQA), The Third Generation ITU-T Standard for End-to-End Speech Quality Measurement Part I—Temporal Alignment,” Journal of the Audio Engineering Society **61**(6), 366–384.

Biberger, T., and Ewert, S. D. (**2016**). “Envelope and intensity based prediction of psychoacoustic masking and speech intelligibility,” The Journal of the Acoustical Society of America **140**(2), 1023–1038, doi: `10.1121/1.4960574`.

Biberger, T., and Ewert, S. D. (**2017**). “The role of short-time intensity and envelope power for speech intelligibility and psychoacoustic masking,” The Journal of the Acoustical Society of America **142**(2), 1098–1111, doi: `10.1121/1.4999059`.

Biberger, T., and Ewert, S. D. (**2022**). “Towards a simplified and generalized monaural and binaural auditory model for psychoacoustics and speech intelligibility,” Acta Acustica **6**, 23, doi: `10.1051/aacus/2022018`.

Biberger, T., Fleßner, J.-H., Huber, R., and Ewert, S. (**2018**). “An Objective Audio Quality Measure Based on Power and Envelope Power Cues,” Journal of the Audio Engineering Society **66**(7/8), 578–593, doi: `10.17743/jaes.2018.0031`.

Biberger, T., Schepker, H., Denk, F., and Ewert, S. D. (**2021**). “Instrumental Quality Predictions and Analysis of Auditory Cues for Algorithms in Modern Headphone Technology,” Trends in Hearing **25**, 233121652110012, doi: `10.1177/23312165211001219`.

Blauert, J. (**1996**). "Spatial Hearing: The Psychophysics of Human Sound Localization," doi: `10.7551/mitpress/6391.001.0001`.

Bronkhorst, and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," The Journal of the Acoustical Society of America **83**(4), 1508–1516, doi: `10.1121/1.395906`.

Bronkhorst, A. W. (**2000**). "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions," Acta Acustica united with Acustica **86**(1), 117–128.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," The Journal of the Acoustical Society of America **92**(6), 3132–3139, doi: `10.1121/1.404209`.

Derleth, P., Georganti, E., Latzel, M., Courtois, G., Hofbauer, M., Raether, J., and Kuehnel, V. (**2021**). "Binaural Signal Processing in Hearing Aids," Seminars in Hearing **42**(3), 206–223, doi: `10.1055/s-0041-1735176`.

Dietz, M., Encke, J., Bracklo, K. I., and Ewert, S. D. (**2021**). "Tone detection thresholds in interaurally delayed noise of different bandwidths," Acta Acustica **5**, 60, doi: `10.1051/aacus/2021054`.

Dietz, M., Ewert, S. D., and Hohmann, V. (**2011**). "Auditory model based direction estimation of concurrent speakers from binaural signals," Speech Communication **53**(5), 592–605, doi: `10.1016/j.specom.2010.05.006`.

Doclo, S., Gannot, S., Moonen, M., and Spriet, A. (**2010**). "Acoustic Beamforming for Hearing Aid Applications," in *Handbook on Array Processing and Sensor Networks*, edited by S. Haykin and K. J. R. Liu (John Wiley & Sons, Inc., Hoboken, NJ, USA), pp. 269–302, doi: `10.1002/9780470487068.ch9`.

Encke, J., and Dietz, M. (**2022**). "A hemispheric two-channel code accounts for binaural unmasking in humans," Communications Biology **5**(1), 1122, doi: `10.1038/s42003-022-04098-x`.

Eurich, B., and Dietz, M. (**2023**). "Fast binaural processing but sluggish masker representation reconfiguration," The Journal of the Acoustical Society of America **154**(3), 1862–1870, doi: `10.1121/10.0021072`.

Eurich, B., Encke, J., Ewert, S. D., and Dietz, M. (**2022**). "Lower interaural coherence in off-signal bands impairs binaural detection," The Journal of the Acoustical Society of America **151**(6), 3927–3936, doi: `10.1121/10.0011673`.

Fleßner, J.-H., Biberger, T., and Ewert, S. D. (**2019**). "Subjective and Objective Assessment of Monaural and Binaural Aspects of Audio Quality," IEEE/ACM Transactions on Audio, Speech, and Language Processing **27**(7), 1112–1125, doi: `10.1109/TASLP.2019.2904850`.

Flessner, J.-H., Huber, R., and Ewert, S. (**2017**). "Assessment and Prediction of Binaural Aspects of Audio Quality," Journal of the Audio Engineering Society **65**(11), 929–942, doi: `10.17743/jaes.2017.0037`.

Friedman, J. H. (**1991**). "Multivariate Adaptive Regression Splines," The Annals of Statistics **19**(1), 1–67, doi: `10.1214/aos/1176347963`.

Gabrielsson, A., and Lindström, B. (**1985**). "Perceived Sound Quality of High-Fidelity Loudspeakers," Journal of the Audio Engineering Society **33**(1/2), 33–53.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hearing Research **47**(1-2), 103–138, doi: `10.1016/0378-5955(90)90170-T`.

Gößling, N., Marquardt, D., and Doclo, S. (**2021**). "Performance Analysis of the Extended Binaural MVDR Beamformer With Partial Noise Estimation," IEEE/ACM Transactions on Audio, Speech, and Language Processing **29**, 462–476, doi: `10.1109/TASLP.2020.3043674`.

Grothe, B., Pecka, M., and McAlpine, D. (**2010**). "Mechanisms of Sound Localization in Mammals," Physiological Reviews **90**(3), 983–1012, doi: `10.1152/physrev.00026.2009`.

Gupta, R., Ranjan, R., He, J., Gan, W.-S., and Peksi, S. (**2020**). "Acoustic transparency in hearables for augmented reality audio: Hear-through techniques review and challenges," in *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society.

Harlander, N., Huber, R., and Ewert, S. D. (**2014**). "Sound Quality Assessment Using Auditory Models," Journal of the Audio Engineering Society **62**(5), 324–336.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," The Journal of the Acoustical Society of America **115**(2), 833–843, doi: `10.1121/1.1639908`.

Heil, P., and Friedrich, B. (**2023**). "How to define thresholds for level and interaural-level-difference discrimination: Insights from scedasticities and distributions," Hearing Research **436**, 108837, doi: `10.1016/j.heares.2023.108837`.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a Gammatone filterbank," Acta Acustica united with Acustica **88**(3), 433–442.

Holdsworth, J., Patterson, R., Nimmo-Smith, I., and Rice, P. (**1988**). "Implementing a Gammatone Filter Bank," Annex C of the SVOS Final Report: Part A: The Auditory Filterbank 1 (1), 1–5.

Huber, R., and Kollmeier, B. (**2006**). "PEMO-Q—A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception," IEEE Transactions on Audio, Speech, and Language Processing **14**(6), 1902–1911, doi: `10.1109/TASL.2006.883259`.

Jekabsons, G. (**2011**). "ARESLab: Adaptive regression splines toolbox for Matlab/Octave" .

Jeub, M., Schafer, M., Esch, T., and Vary, P. (**2010**). "Model-Based Dereverberation Preserving Binaural Cues," IEEE Transactions on Audio, Speech, and Language Processing **18**(7), 1732–1745, doi: `10.1109/TASL.2010.2052156`.

Kates, J. M., and Arehart, K. H. (**2010**). “The Hearing-Aid Speech Quality Index (HASQI),” Journal of the Audio Engineering Society **58**(5), 363–381.

Kates, J. M., and Arehart, K. H. (**2014**). “The Hearing-Aid Speech Quality Index (HASQI) Version 2,” Journal of the Audio Engineering Society **62**(3), 99–117.

Kates, J. M., and Arehart, K. H. (**2016**). “The Hearing-Aid Audio Quality Index (HAAQI),” IEEE/ACM Transactions on Audio, Speech, and Language Processing **24**(2), 354–365, doi: `10.1109/TASLP.2015.2507858`.

Klug, J., and Dietz, M. (**2022**). “Frequency dependence of sensitivity to interaural phase differences in pure tones,” The Journal of the Acoustical Society of America **152**(6), 3130–3141, doi: `10.1121/10.0015246`.

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). “The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers,” The Journal of the Acoustical Society of America **108**(2), 723–734, doi: `10.1121/1.429605`.

Kollmeier, B., and Peissig, J. (**1990**). “Speech Intelligibility Enhancement by Interaural Magnification,” Acta Oto-Laryngologica **109**(sup469), 215–223, doi: `10.1080/00016489.1990.12088432`.

Majdak, P., Hollomey, C., and Baumgartner, R. (**2022**). “AMT 1.x: A toolbox for reproducible research in auditory modeling,” Acta Acustica **6**, 19, doi: `10.1051/aacus/2022011`.

Manocha, P., Kumar, A., Xu, B., Menon, A., Gebru, I. D., Ithapu, V. K., and Calamia, P. (**2022**). “SAQAM: Spatial Audio Quality Assessment Metric,” in *Interspeech 2022*, ISCA, pp. 649–653, doi: `10.21437/Interspeech.2022-406`.

Marquardt, D., Hohmann, V., and Doclo, S. (**2015**). “Interaural Coherence Preservation in Multi-Channel Wiener Filtering-Based Noise Reduction for Binaural Hearing Aids,” IEEE/ACM Transactions on Audio, Speech, and Language Processing **23**(12), 2162–2176, doi: `10.1109/TASLP.2015.2471096`.

Mc Laughlin, M., Franken, T. P., van der Heijden, M., and Joris, P. X. (**2014**). “The Interaural Time Difference Pathway: A Comparison of Spectral Bandwidth and Correlation Sensitivity at Three Anatomical Levels,” Journal of the Association for Research in Otolaryngology **15**(2), 203–218, doi: `10.1007/s10162-013-0436-6`.

McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). “A neural code for low-frequency sound localization in mammals,” Nature Neuroscience **4**(4), 396–401, doi: `10.1038/86049`.

Moore, B. C. J., and Tan, C.-T. (**2004**). “Development and Validation of a Method for Predicting the Perceived Naturalness of Sounds Subjected to Spectral Distortion,” Journal of the Audio Engineering Society **52**(9), 900–914.

Moore, B. C. J., Tan, C.-T., Zacharov, N., and Mattila, V.-V. (**2004**). “Measuring and Predicting the Perceived Quality of Music and Speech Subjected to Combined Linear and Nonlinear Distortion,” Journal of the Audio Engineering Society **52**(12), 1228–1244.

Munson, W. A., and Gardner, M. B. (**2005**). "Standardizing Auditory Tests," The Journal of the Acoustical Society of America **22**(5_Supplement), 675, doi: `10.1121/1.1917190`.

Nordholm, S., Schepker, H., Tran, L. T. T., and Doclo, S. (**2018**). "Stability-controlled hybrid adaptive feedback cancellation scheme for hearing aids," The Journal of the Acoustical Society of America **143**(1), 150–166, doi: `10.1121/1.5020269`.

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). "An efficient auditory filterbank based on the gammatone function," in *Paper Presented at a Meeting of the IOC Speech Group on Auditory Modelling at RSRE.*

Qiao, Y., Zacharov, N., and Hoffmann, P. F. (**2022**). "Prediction of timbral, spatial, and overall audio quality with independent auditory feature mapping," in *Audio Engineering Society Convention 153*, Audio Engineering Society.

Remme, M. W. H., Donato, R., Mikiel-Hunter, J., Ballestero, J. A., Foster, S., Rinzel, J., and McAlpine, D. (**2014**). "Subthreshold resonance properties contribute to the efficient coding of auditory spatial cues," Proceedings of the National Academy of Sciences **111**(22), E2339–E2348, doi: `10.1073/pnas.1316216111`.

Rennies, J., and Kidd, G. (**2018**). "Benefit of binaural listening as revealed by speech intelligibility and listening effort," The Journal of the Acoustical Society of America **144**(4), 2147–2159, doi: `10.1121/1.5057114`.

Rix, A., Beerends, J., Hollier, M., and Hekstra, A. (**2001**). "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, Vol. 2, pp. 749–752 vol.2, doi: `10.1109/ICASSP.2001.941023`.

Rohdenburg, T., Hohmann, V., and Kollmeier, B. (**2007**). "Robustness Analysis of Binaural Hearing Aid Beamformer Algorithms by Means of Objective Perceptual Quality Measures," in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 315–318, doi: `10.1109/ASPAA.2007.4393016`.

Rumsey, F., Zieliński, S., Kassier, R., and Bech, S. (**2005**). "On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality," The Journal of the Acoustical Society of America **118**(2), 968–976, doi: `10.1121/1.1945368`.

Schäfer, M., Bahram, M., and Vary, P. (**2013**). "An extension of the PEAQ measure by a binaural hearing model," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8164–8168, doi: `10.1109/ICASSP.2013.6639256`.

Schepker, H., Denk, F., Kollmeier, B., and Doclo, S. (**2019**). "Subjective Sound Quality Evaluation of an Acoustically Transparent Hearing Device," in *Audio Engineering Society Conference: 2019 AES International Conference on Headphone Technology*, Audio Engineering Society.

Schepker, H., Denk, F., Kollmeier, B., and Doclo, S. (**2020**). "Acoustic Transparency in Hearables—Perceptual Sound Quality Evaluations," Journal of the Audio Engineering Society **68**(7/8), 495–507.

Seo, J.-H., Chon, S. B., Sung, K.-M., and Choi, I. (**2013**). “Perceptual Objective Quality Evaluation Method for High Quality Multichannel Audio Codecs,” Journal of the Audio Engineering Society **61**(7/8), 535–545.

Series, B. (**2014**). “Method for the subjective assessment of intermediate quality level of audio systems,” International Telecommunication Union Radiocommunication Assembly .

Takanen, M., Santala, O., and Pulkki, V. (**2014**). “Visualization of functional count-comparison-based binaural auditory model output,” Hearing Research **309**, 147–163, doi: `10.1016/j.heares.2013.10.004`.

Temme, S. F. (**2019**). “Testing Audio Performance of Hearables,” in *Audio Engineering Society Conference: 2019 AES International Conference on Headphone Technology*, Audio Engineering Society.

Thiemann, J., Müller, M., Marquardt, D., Doclo, S., and van de Par, S. (**2016**). “Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene,” EURASIP Journal on Advances in Signal Processing **2016**(1), 12, doi: `10.1186/s13634-016-0314-6`.

Toole, F. E. (**1985**). “Subjective Measurements of Loudspeaker Sound Quality and Listener Performance,” Journal of the Audio Engineering Society **33**(1/2), 2–32.

Yousefian, N., Loizou, P. C., and Hansen, J. H. (**2014**). “A coherence-based noise reduction algorithm for binaural hearing aids,” Speech Communication **58**, 101–110, doi: `10.1016/j.specom.2013.11.003`.

CHAPTER 6

General discussion and conclusion

Three models of effective binaural processing have been presented, each of them based on the complex correlation coefficient γ. Two of them are basic research and have addressed apparent contradictions on the analysis window size in binaural compared to monaural hearing (chapters 3 and 4). The third model transferred γ to the engineering field (chapter 5) by incorporating γ into a computationally efficient measure of monaural and binaural audio quality, targeted at developers of algorithms in hearing technology (termed eMoBi-Q). In the following, the scientific contributions are evaluated on a higher level. First, the proposed concept of interference across frequency and time will be contextualized (section 6.1), then the insights gained from the use of γ are considered (section 6.2). Finally, the impact of the presented findings are summarized and implications for future research are suggested (sections 6.3 and 6.4).

6.1 Spectral and temporal interference of IPD statistics

Some experimental results suggested larger binaural than monaural filter bandwidths (van der Heijden and Trahiotis, 1999; Holube *et al.*, 1998; Kolarik and Culling, 2010) or time windows (Kollmeier and Gilkey, 1990; Grantham and Wightman, 1979; Holube *et al.*, 1998). However, other results can be best or even only explained assuming that binaural hearing exploits the full spectral (Bernstein and Trahiotis, 2020b; Dietz *et al.*, 2021; Langford and Jeffress, 1964; Rabiner *et al.*, 1966) and temporal (Siveke *et al.*, 2008; Dietz *et al.*, 2008) resolution provided by the basilar membrane filters [as estimated by Glasberg and Moore (1990)]. Assuming interference mechanisms as introduced in section 2.2 instead of overall larger analysis windows solved two long-lasting conceptual contradictions of binaural research.

The models presented in the chapters 3 and 4 each account for both, the results suggesting a generally lower resolution and those suggesting access to the full resolution provided by the basilar membrane. The working hypothesis of this thesis was that the binaural system has the same high resolution as the monaural system. That means there is no hard-wired convergence or averaging in the spectral or temporal encoding of sound. This is based on the considerations that a higher-resolution system might potentially be unable able to access the full resolution in certain cases. However, a low-resolution system cannot provide high resolution in certain cases. Our hypothesis is in line with the ability to detect auditory events that require the full frequency selectivity and temporal resolution provided by the basilar membrane. However, under certain circumstances, this resolution cannot be fully accessed. Such hypothesized reductions in resolution are associated with auditory processing beyond peripheral encoding. Auditory events are perceptually organized and interpreted (Bregman, 1990). They are compared and perceived in context with previous or ongoing events as well as with expectation (Sutojo *et al.*, 2020). Based on the similarity of extracted attributes, summarized as Gestalt rules (Wertheimer, 1923), auditory objects are formed. They group and separate sound components across both frequency and time (Middlebrooks and Simon, 2017). Efficiency is increased by predictive coding and updating (Francis and Wonham, 1976; Majdak *et al.*, 2020). The more abstract a perceptual representation of a sound event, the larger the corresponding analysis window (Bizley and Cohen, 2013; Simon, 2017; Elhilali, 2017): While auditory periphery provides a high frequency selectivity and temporal resolution, more central stages form more holistic representations (see Fig. 6.1).

Thus, the properties of all auditory pathway stages have an impact on whether a sound is perceived. Therefore it is unsurprising that the detectability of a certain

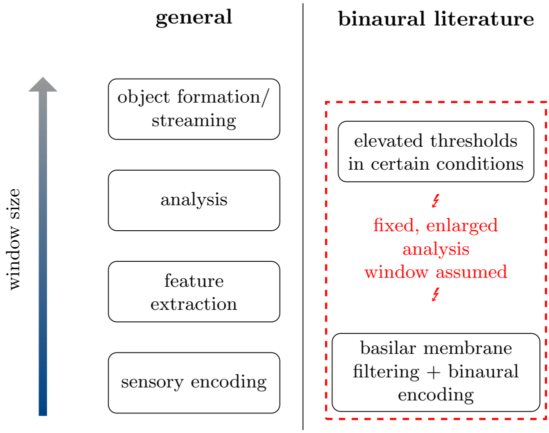
general
binaural literature
window size
object formation/ streaming
analysis
feature extraction
sensory encoding
elevated thresholds in certain conditions
↯
fixed, enlarged analysis window assumed
↯
basilar membrane filtering + binaural encoding

Figure 6.1: Schematic illustration of the discussed discrepancy between general principles of neurosensory processing (left column) and the larger analysis windows reported in binaural literature (right column).

sound attribute depends on the spectral and temporal context. Adding a pure tone to a noise that differs in its IPD statistics leads to a modification of the resulting interaural parameters in a given frequency band or at a given time. However, if similar changes in IPD statistics occur in spectral or temporal proximity, it is reasonable to assume that tone detection will be affected. Although top-down attentional processes play a major role in shaping auditory perception (Shinn-Cunningham *et al.*, 2017), the proposed models (Eurich *et al.*, 2022; Eurich and Dietz, 2023) account for the apparently larger binaural than monaural analysis windows only involving bottom-up processes. In the spectral domain, incoherence interference from off-signal bands was introduced. In the temporal domain, across-time interference was implemented as a "sluggish" reformation of reference IPD statistics. The presented effective, phenomenological models did not implement a clear neurophysiologic or neurocognitive mechanism causing the spectral and temporal interference. Instead, the origins are associated with auditory object formation. In the spectral domain, the superposition of two opposingly delayed noises produces frequency regions with different interaural correlation. This can be perceived as different objects, distinguished by their spectra. Additionally, two opposingly lateralized noise components can be perceived. In case of a single delayed noise, however, only one lateralized noise component is perceived, constant in correlation across frequency. It is hypothesized that detection of the widening cue (i.e. a less compact within-the-head representation, see section 2.1) induced by the tone is impaired by the more complex internal representation (which means a model of the interaural statistics, see chapter 4) of the masker, involving more auditory objects. Similar effects have been described as binaural interference (Bernstein and Trahiotis, 1995) and modulation detection interference (Yost and Sheft, 1989; Bacon and Konrad, 1993; Mendoza *et al.*, 1995; Oxenham and Dau, 2001). In the temporal domain, the changing masker IPD statistics are hypothesized to entail reformation of the masker auditory object which tone detection relies on. That means the perceptual distance between the internal representations of target and masker converges in a "sluggish" way. However, it is possible that the proposed interference mechanisms rather result from processes prior to or interacting with object formation, however, beyond the binaural brainstem.

While the proposed explanations to the apparent "wider binaural filters" and "binaural sluggishness" follow the same idea, the histories of those assumptions differ:

There have been simultaneous, opposing statements on the bandwidth determining binaural detection: from the increase of detection thresholds as a function of masker ITD, Rabiner *et al.* (1966) derived the bandwidth underlying binaural detection to be very similar to the bandwidth underlying monaural detection. Sondhi and Guttman (1966) derived a larger bandwidth for binaural detection compared to that known from monaural detection, based on elevated thresholds when the masker IPD changed in spectral proximity to the target. However, for modeling binaural detection, either the underlying bandwidth added a degree of freedom [van der Heijden and Trahiotis (1999) vs. Bernstein and Trahiotis (2020a)], or conditions highlighting the conceptual difference [most importantly: impaired detection in opposingly delayed noises (van der Heijden and Trahiotis, 1999)] did not receive appropriate attention (Bernstein and Trahiotis, 2015, 2017, 2020a; Breebaart *et al.*, 2001b). Dietz *et al.* (2021) provided further evidence for the conventional critical bandwidth accounting for binaural detection, Marquardt and McAlpine (2009) and Eurich *et al.* (2022, i.e. chapter 3) provided a conceptual and quantitative explanation for the bandwidth appearing larger in certain conditions.

On the other hand, more "sluggish" binaural than monaural processing has been accepted since the late 1970s. This "sluggishness" assumption was based on data showing elevated thresholds in temporal proximity to changes in masker IPD statistics (Grantham and Wightman, 1978, 1979). Thus, most models have assumed fixed temporal integration [e.g., Breebaart *et al.* (2001c); Hauth and Brand (2018)]. Evidence that binaural processing speed is only limited by peripheral filtering was presented thirty years later (Siveke *et al.*, 2008; Dietz *et al.*, 2008). Chapter 4 helps unifying fast and "sluggish" processing and therefore adds to a previously limited scientific debate. Fast binaural processing is also supported by Bischof *et al.* (2023), who found that the altered binaural unmasking of speech in the presence of room reflections and late reverberation is better explained by 12 ms than by 300 ms windows.

Notwithstanding the explanation for the sometimes apparently larger binaural than monaural analysis windows given in chapters 3 and 4, integration in the sense of optimal combination is assumed as a subsequent stage. As pointed out and modeled, it is reasonable to assume (1) an optimal combination of frequency bands, accounting for off-frequency exploitation in narrow-band maskers (Hall *et al.*, 1983; van de Par and Kohlrausch, 1999; Breebaart *et al.*, 2001b), and (2) an optimal combination of time segments to account for temporal integration and the multiple-looks hypothesis (Viemeister and Wakefield, 1989). Both is in line with auditory object formation operating on integrated spectrotemporal information (Hsieh *et al.*, 2018). This is accounted for in the presented models by locating the interference mechanism in the frontend while locating integration in the backend (see Table 6.1) . Although associated with non-ignorable off-signal changes on a cortical level, it is thought separate from even higher-level integration, let alone attention-driven processes (not modeled). In summary, the proposed work has established that detrimental interference helps to reconcile the requirements of high and low binaural resolution.

6.2 Demonstrating the viability of the two-channel code for effective binaural modeling

6.2.1 The two-channel concept catches on

In addition to establishing the interference concept discussed above, the models proposed in this thesis also continue the line of arguments regarding the viability of the two-channel code as an effective explanation of MSO processing.

As considered in chapter 2, modeling binaural unmasking was dominated by an approach based on the delay-line concept first stated by (Jeffress, 1948). The comprehensiveness and accuracy remained unmatched for a long time. Breebaart *et al.* (2001a) combined the delay-line concept with contralaterial inhibition of the ipsilateral signal, enabling the model also to account for ILD processing. A variety of binaural psychophysic results could be reproduced (Breebaart *et al.*, 2001b,c). As a further step to account for more recent insights of mammals' MSO processing, Dietz *et al.* (2008) incorporated the excitatory-inhibitory concept by Breebaart *et al.* (2001a) into a framework that extracted IPDs instead of time-delay-compensating coincidence detection, based on evidence for phase-specific rate coding (Brand *et al.*, 2002; Marquardt and Mcalpine, 2007). As a more appropriate description of MSO processing in mammals, a hemispheric two-channel code was suggested (Grothe *et al.*, 2010; McAlpine *et al.*, 2001). In the following, Encke and Dietz (2022) interpreted the MSO two-channel code as two orthogonal correlation coefficients and expressed this as the complex correlation coefficient γ. Similar to the single-correlation-coefficient-based approach by Bernstein and Trahiotis (2017), Encke and Dietz (2022) accounted for a variety of binaural detection thresholds including their dependence of masker ITD and correlation. Both models involved a single-ERB gammatone filter and two degrees of freedom for the binaural model. As the γ feature used by Encke and Dietz (2022) additionally encodes the IPD of the stimulus, the periodically oscillating thresholds as a function of masker ITD was also captured, which would have required the approach by Bernstein and Trahiotis (2017) to evaluate a larger range of the correlation function. As the usage of the IPD is more likely to reflect mammals' binaural processing than extracting interaural delays of several milliseconds, Encke and Dietz (2022) reached a new level of predictive power paired with physiological plausibility. However, the experimental condition that really highlights the conceptual differences between model concepts that do or do not involve compensation of large masker delays, is the impaired 500-Hz tone detection in the presence of two opposingly delayed noises. This condition was previously only addressed by van der Heijden and Trahiotis (1999) and Marquardt and McAlpine (2009). Only the delay-line model of van der Heijden and Trahiotis (1999) satisfactorily explained the data, thus it was taken as proof of the delay-line concept as an effective binaural model. Since both Bernstein and Trahiotis (2017) and Encke and Dietz (2022) used only one single-ERB gammatone filter, neither model could have reproduced all threshold elevations resulting from adding a masker with an ITD opposite to that of the first masker. As discussed in chapter 3, at masker ITDs that correspond to an on-frequency IPD of zero, this leads to a threshold elevation even though the coherence is only affected in off-signal bands. A filter bandwidth larger than the conventional ERB $= 79$ Hz can thus account for the higher threshold. While a delay-line has no effect in opposing delays, it has in the single-delay case. Thus, it reduces the threshold in case of one delayed masker. Since the latter threshold is determined by a conventional filter bandwidth anyway (see section 2.1.3)and the delay compensation itself is questionable, chapter 3 replaced the larger filter plus delay-line by a conventional filter plus across-frequency incoherence interference, an evolution of the hypothesis proposed by Marquardt and McAlpine (2009). That is, two debatable assumptions were removed and one new assumption was introduced. The plausibility of the new assumption, i.e. applying the interference concept which is a core concept of this thesis, was discussed in section 6.1. Explaining the apparent delay-line proof without the delay-line, one critical assumption less, and one arguably more plausible new assumption is proposed as the final step in showing that the two-channel code is the more viable concept for MSO processing than the delay-line.

6.2.2 The potential of γ as a new workhorse

As introduced in chapter 2, the complex correlation coefficient γ was introduced to binaural research by Encke and Dietz (2022) as a mathematically efficient description of the physiologally plausible hemispheric two-channel code. It was taken up in all three models presented in this thesis. Table 6.1 summarizes the architectures of the presented models: γ was used as a single-channel (chapter 4), single-channel with off-signal interference (chapter 3) and multi-channel (chapter 5) feature as well as with a single-frame (chapter 3), with consecutive frames (chapter 5) and with instantenous extraction (chapter 4). In case of multi-channel backends, time or frequency elements were optimally combined, i.e. $d' = \sqrt{\sum_n {d'_n}^2}$, mimicking linear independent time or frequency elements. Furthermore, in chapter 5, for channels with center frequencies above 1300 Hz, γ was computed from the envelope of the band-pass filtered signals while channels at lower frequencies used the TFS (roll-off frequency: 1300 Hz). This reflects the sensitivity of the MSO and LSO to interaural disparities in the TFS and the envelope, respectively (Remme *et al.*, 2014; Klug and Dietz, 2022). The high predictive power across the whole range of use cases, stimuli and tasks demonstrates that γ is a viable and well-suited feature for effective binaural models. Its viability was demonstrated by its accurate representation of binaural psychophysics, as well as its good reproduction of subjective ratings of binaural audio quality, atypically without any nonlinear back-end decoding.

		Eurich *et al.* (2022)	**Eurich and Dietz (2023)**	**eMoBi-Q**
frontend	frequency	on-signal + interference	on-signal	multiple
	time	average	instantaneous + interference	multiple
	features	γ	γ	γ (roll-off 1300 Hz)
		DC power	DC power	DC power
				ILD
backend	frequency	single	single	multiple
	time	single	multiple looks	multiple
focus		tone detection spectral context	tone detection temporal context	audio quality

Table 6.1: Overview over the three presented models on their architecture concerning spectral and temporal processing as well as the focus. Green: Single analysis window, i.e. frequency band or time frame; on-signal denotes the frequency band centered at the tone frequency. orange: multiple analysis windows, i.e. frequency bands or time frames. red: interference across frequency or time. In multi-channel- or multi-frame backends, time or frequency elements were combined in an optimal manner.

However, there are limitations: The formulation of the two-channel code as a complex number entails a circular complex plane as resulting from sinusoidal IPD-rate functions with best IPDs differing by $\frac{\pi}{2}$, e.g., $\pm\frac{\pi}{4}$ (see chapter 2). This has the consequence that γ codes a change in IPD irrespective of the reference IPD while experimental results show a lower sensitivity to IPDs around $\pm\pi$ than around 0 [apparent in the detection thresholds by, e.g., Yost (1974); van der Heijden and Trahiotis (1999); van de Par and Kohlrausch (1999)]. Furthermore, the assumption of sinusoidal IPD-rate functions is a simplificiation that neglects hair-cell processing, i.e. demodulation through half-wave rectification (HWR) and low-pass filtering. Harmonics added by HWR combined with attenuation of the higher harmonics through low-pass filtering results in a frequency-dependent modification of the rate-IPD functions and thus of the feature space. The precise modeling results based on a single 500 Hz-centered channel in Encke and Dietz (2022) and Eurich *et al.*

(2022, chapter 3) have shown that the simplification of only assuming correlator units at $\pm\frac{\pi}{4}$ is acceptable for low frequencies. However, by not taking the reference-dependent IPD sensitivity into account, the threshold difference between N_0S_π and $N_\pi S_0$ was not reproduced but required adjusting the internal noise parameter. The models presented in this thesis do not involve hair-cell processing anyway. Therefore, they point out the shortcomings resulting from the lack of hair-cell processing but circumvent the violation of assumptions underlying γ. Encke and Dietz (2022) discussed a modification of γ in order to reduce IPD sensitivity at higher reference IPDs. For a more comprehensive approach, however, an implementation of MSO encoding based on nonlinear peripheral processing should be considered in future work.

6.3 Impact of the findings

It was discussed above that the new concepts applied and presented in this thesis provide evidence for new perspectives on binaural processing. Namely, (1) the interference concept with its two applications in the spectral and temporal domains help unifying apparent conceptual contradictions in binaural research, and (2) the hemispheric two-channel code is the conceptually more consistent model of MSO processing than the delay line. However, the quantitative consequences and benefits of the new approaches need to be considered in a nuanced way. In psychoacoustics, conditions have been designed that clearly point out the different requirements on binaural analysis windows [spectral: van der Heijden and Trahiotis (1999), Holube *et al.* (1998), Dietz *et al.* (2021), Bernstein and Trahiotis (2020b); temporal: Grantham and Wightman (1978), Holube *et al.* (1998), Siveke *et al.* (2008)]. The conditions that demonstrate the plausibility of a filter bandwidth of ERB = 79 Hz (for a 500 Hz center frequency) in combination with an additional detrimental impact are the detection thresholds for S_π in a single noise versus those in two opposingly delayed noises with each ITD = 2 ms (And, similarly, S_0 detection thresholds at ITD = 1 ms). These thresholds differ by about 4 dB (van der Heijden and Trahiotis, 1999)[1].

---

[1] As pointed out in section 2.1.3, the filter bandwidth dictates the maximum binaural unmasking in delayed noise. For ITDs above 4 ms, (Dietz *et al.*, 2021) showed that filter bandwidths of ERB $\leq$ 79 Hz best explains the noise-bandwidth dependent masking pattern, based on a measure of IPD variability, for this usecase comparable to the coherence $|\gamma|$. For lower ITDs, however, the predicted thresholds for ERB = 79 Hz and ERB = 130 Hz differed by only about 1 dB. Thus, mathematical and conceptual consistency sometimes has nuanced quantitative consequences.

Although standard errors are larger in the opposingly-delayed case, the difference is clear and could not be reproduced with one single-ERB channel. Therefore, accounting for such clear and decisive psychoacoustic conditions requires models to provide a unifying concept, such as the proposed across-frequency incoherence interference. In complex sounds like speech and music as well as free-field complex acoustic conditions involving distortions and reverberation, however, interaural differences are reduced by the stronger spectrotemporal fluctuations, redundancy and diffuseness (Bronkhorst and Plomp, 1988; Bronkhorst, 2000; Zahorik, 2021). Thus, the binaural advantage is reduced (Culling *et al.*, 1994; George *et al.*, 2012; Beutelmann and Brand, 2006; Biberger and Ewert, 2022). This is hypothesized to limit the implication of the across-frequency incoherence interference, although decisive in the "laboratory" tone-in-noise conditions. Consequently, in the eMoBi-Q model, developed to transfer the psychoacoustically validated γ feature to the application of audio quality assessment (chapter 5), across-frequency incoherence interference was not included as it did not improve performance. With the focus being computational efficiency, consecutive frames of 400 ms were used instead of the instantaneous feature extraction suggested in chapter 4. This approach is therefore not expected to account for psychoacoustic results regarding fast binaural processing modeled in chapter 4, such as phasewarp detection (Siveke *et al.*, 2008) or for sound source localization based on short glimpses (Dietz *et al.*, 2011). However, sound quality of speech and music signals, processed by algorithms as typically applied in modern hearing devices, are well captured by the consecutive 400 ms used in eMoBi-Q. This underlines that the perceptual relevance of the maximum temporal binaural resolution depends on the use case. Longer-term measures account for sound quality in modern hearing technology as well as tone detection and speech intelligility in stationary maskers [e.g., Eurich *et al.* (2022) and Beutelmann and Brand (2006), respectively]. For detection of rapid fluctuations of speech intelligibility in modulated maskers, however, exploitation of the full temporal resolution provided by the basilar membrane is required [e.g., Eurich and Dietz (2023) and Beutelmann *et al.* (2010)]. In a nutshell, this thesis provides conceptual "game changers". At the same time it demonstrates that the implications for application-oriented model use in real-world acoustic scenarios are subtle.

6.4 Implications for future research

As discussed above, the concept of IPD statistics interfering across frequency provides the missing milestone in favour of the two-channel code to be the more plausible and more consistent binaural model concept than the long dominating delay line. At the same time, it provided an explanation for the sometimes apparently larger binaural than monaural filter bandwidth. Furthermore, applying the concept of IPD statistics interfering across time provided an explanation to the apparent contradiction of fast versus “sluggish” binaural processing. The implication for future research is that the binaural processing speed can be assumed to be limited only by basilar-membrane filtering. Non-ignorable off-signal IPD statistics should be considered as a reason in case of discrepant observations. For modeling effective binaural processing, the two-channel concept may be the basis, assuming the same processing bandwidth and speed for monaural and binaural processing. Depending on the specific use case, interference mechanisms may be beneficial.

The presented research further suggests that the complex correlation coefficient γ as introduced to binaural modeling by Encke and Dietz (2022) is suitable for further employment in binaural models and algorithms aiming to describe the effective IPD/ITD encoding. The waveform-based model (Eurich *et al.*, 2022), incorporating γ, is publicly available through the AMT (Majdak *et al.*, 2022) for free usage and further development. Future research could assess the implications of cochlear hearing loss on binaural hearing when combining γ with a nonlinear model of cochlear processing that can be modified to simulate hair cell loss [e.g., the gammawarp filterbank (Kates and Prabhu, 2018)]. The combined perceptual validity and computational efficiency of γ make it suitable for individual model-based diagnostics [like, e.g., Herrmann and Dietz (2021)], frameworks on computational auditory scene analysis [e.g., aiming at sound source separation (Sutojo *et al.*, 2022; Kong *et al.*, 2020)] and real-time simulations of hearing-impaired performance [like, e.g., (Grimault *et al.*, 2016)]. Moreover, backends such as neural networks trained on the outputs of γ, an LSO-inspired feature and a monaural feature, appear useful in addressing tasks that require more sophisticated central processing. Machine hearing algorithms with a focus on computational efficiency could also benefit from γ as it provides a measure that reflects IPD statistics by calculating only two correlation coefficients. Also backends for direction-of-arrival (DOA) estimation of sound sources could be designed based on γ. However, various successful approaches on DOA estimation are available [Bayesian inference: Barumerli *et al.* (2020); geometretic model: Barot *et al.* (2023); deep neural network at output of peripheral model: Goli and van de Par (2023), Hermitian angle spectrum of relative transfer function vectors between signal and reference: Fejgin and Doclo (2022)]. Addressing the overarching challenges of competing sources, interference and reverberation is itself a highly complex research topic (see May *et al.* 2013 for review), where recent approaches apply machine-learning based methods (Örnolfsson *et al.*, 2021; Ren *et al.*, 2021; Wu *et al.*, 2021).

The proposed model for the combined assessment of monaural and binaural audio quality (eMoBi-Q, chapter 5) can form the basis for real-time control of hearing aid algorithms. Combining the γ feature with local DC power and ILDs has been shown to reflect audio quality to a high degree. Algorithm developers in both science and engineering can also use the features of eMoBi-Q as a monitoring tool, benefiting from their perceptual validity coupled with simplicity.

6.5 Conclusion

A unifying solution has been suggested for the previously contradicting statements on the spectral and temporal resolution of binaural hearing. For both domains, a high binaural resolution combined with an interference concept has been shown to unify apparently contradicting experimental results. At the same time, the physiologically plausible concept of a two-hemispheric-channel code has been shown to be also the more consistent model than the concept of internal compensation of interaural delays. This is contributing decisively to the replacement of one of the oldest and most dominant models of sensory neuroscience. The complex correlation coefficient γ has been shown to be a viable and mathematically efficient measure to binaural perception. At the same time it reflects the two-channel code associated with mammalian binaural processing. To demonstrate its practical applicability, a model has been presented to instrumentally assess audio quality of hearing algorithm and loudspeaker processing.

References

Bacon, S. P., and Konrad, D. L. (**1993**). "Modulation detection interference under conditions favoring within- or across-channel processing," The Journal of the Acoustical Society of America **93**(2), 1012–1022, doi: `10.1121/1.405549`.

Barot, P., Mombaur, K., and MacDonald, E. (**2023**). "Estimating speaker direction on a humanoid robot with binaural acoustic signals" doi: `10.48550/arXiv.2307.12129`.

Barumerli, R., Majdak, P., Reijniers, J., Baumgartner, R., Geronazzo, M., and Avanzini, F. (**2020**). "Predicting Directional Sound-Localization of Human Listeners in both Horizontal and Vertical Dimensions," in *Audio Engineering Society Convention 148*, Audio Engineering Society.

Bernstein, L. R., and Trahiotis, C. (**1995**). "Binaural interference effects measured with masking-level difference and with ITD- and IID-discrimination paradigms," The Journal of the Acoustical Society of America **98**(1), 155–163, doi: `10.1121/1.414467`.

Bernstein, L. R., and Trahiotis, C. (**2015**). "Converging measures of binaural detection yield estimates of precision of coding of interaural temporal disparities," The Journal of the Acoustical Society of America **138**(5), EL474–EL479, doi: `10.1121/1.4935606`.

Bernstein, L. R., and Trahiotis, C. (**2017**). "Binaural detection-based estimates of precision of coding of interaural temporal disparities across center frequency," The Journal of the Acoustical Society of America **141**(5), 3973–3973, doi: `10.1121/1.4989060`.

Bernstein, L. R., and Trahiotis, C. (**2020**a). "Binaural detection as a joint function of masker bandwidth, masker interaural correlation, and interaural time delay: Empirical data and modeling," The Journal of the Acoustical Society of America **148**(6), 3481–3488, doi: `10.1121/10.0002869`.

Bernstein, L. R., and Trahiotis, C. (**2020**b). "A crew of listeners with no more than "slight" hearing loss who exhibit binaural deficits also exhibit higher levels of stimulus-independent internal noise," The Journal of the Acoustical Society of America **147**(5), 3188–3196, doi: `10.1121/10.0001207`.

Beutelmann, R., and Brand, T. (**2006**). “Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners,” The Journal of the Acoustical Society of America **120**(1), 331–342, doi: `10.1121/1.2202888`.

Beutelmann, R., Brand, T., and Kollmeier, B. (**2010**). “Revision, extension, and evaluation of a binaural speech intelligibility model,” The Journal of the Acoustical Society of America **127**(4), 2479–2497, doi: `10.1121/1.3295575`.

Biberger, T., and Ewert, S. D. (**2022**). “Binaural detection thresholds and audio quality of speech and music signals in complex acoustic environments,” Frontiers in Psychology **13**.

Bischof, N. F., Aublin, P. G., and Seeber, B. U. (**2023**). “Fast processing models effects of reflections on binaural unmasking,” Acta Acustica **7**, 11, doi: `10.1051/aacus/2023005`.

Bizley, J. K., and Cohen, Y. E. (**2013**). “The what, where and how of auditory-object perception,” Nature Reviews Neuroscience **14**(10), 693–707, doi: `10.1038/nrn3565`.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). “Precise inhibition is essential for microsecond interaural time difference coding,” Nature **417**(6888), 543–547, doi: `10.1038/417543a`.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**a). “Binaural processing model based on contralateral inhibition. I. Model structure,” The Journal of the Acoustical Society of America **110**(2), 1074–1088, doi: `10.1121/1.1383297`.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**b). “Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters,” The Journal of the Acoustical Society of America **110**(2), 1089–1104, doi: `10.1121/1.1383298`.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**c). “Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters,” The Journal of the Acoustical Society of America **110**(2), 1105–1117, doi: `10.1121/1.1383299`.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (The MIT Press).

Bronkhorst, and Plomp, R. (**1988**). “The effect of head-induced interaural time and level differences on speech intelligibility in noise,” The Journal of the Acoustical Society of America **83**(4), 1508–1516, doi: `10.1121/1.395906`.

Bronkhorst, A. W. (**2000**). “The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions,” Acta Acustica united with Acustica **86**(1), 117–128.

Culling, J. F., Summerfield, Q., and Marshall, D. H. (**1994**). “Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels,” Speech Communication **14**(1), 71–95, doi: `10.1016/0167-6393(94)90058-2`.

Dietz, M., Encke, J., Bracklo, K. I., and Ewert, S. D. (**2021**). "Tone detection thresholds in interaurally delayed noise of different bandwidths," Acta Acustica **5**, 60, doi: `10.1051/aacus/2021054`.

Dietz, M., Ewert, S. D., and Hohmann, V. (**2011**). "Auditory model based direction estimation of concurrent speakers from binaural signals," Speech Communication **53**(5), 592–605, doi: `10.1016/j.specom.2010.05.006`.

Dietz, M., Ewert, S. D., Hohmann, V., and Kollmeier, B. (**2008**). "Coding of temporally fluctuating interaural timing disparities in a binaural processing model based on phase differences," Brain Research **1220**, 234–245, doi: `10.1016/j.brainres.2007.09.026`.

Elhilali, M. (**2017**). "Modeling the Cocktail Party Problem," in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay, **60** (Springer International Publishing, Cham), pp. 111–135, doi: `10.1007/978-3-319-51662-2_5`.

Encke, J., and Dietz, M. (**2022**). "A hemispheric two-channel code accounts for binaural unmasking in humans," Communications Biology **5**(1), 1122, doi: `10.1038/s42003-022-04098-x`.

Eurich, B., and Dietz, M. (**2023**). "Fast binaural processing but sluggish masker representation reconfiguration," The Journal of the Acoustical Society of America **154**(3), 1862–1870, doi: `10.1121/10.0021072`.

Eurich, B., Encke, J., Ewert, S. D., and Dietz, M. (**2022**). "Lower interaural coherence in off-signal bands impairs binaural detection," The Journal of the Acoustical Society of America **151**(6), 3927–3936, doi: `10.1121/10.0011673`.

Fejgin, D., and Doclo, S. (**2022**). "Coherence-Based Frequency Subset Selection for Binaural RTF-vector-based Direction of Arrival Estimation for Multiple Speakers," in *2022 International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 1–5, doi: `10.1109/IWAENC53105.2022.9914768`.

Francis, B. A., and Wonham, W. M. (**1976**). "The internal model principle of control theory," Automatica **12**(5), 457–465, doi: `10.1016/0005-1098(76)90006-6`.

George, E. L. J., Festen, J. M., and Theo Goverts, S. (**2012**). "Effects of reverberation and masker fluctuations on binaural unmasking of speech," The Journal of the Acoustical Society of America **132**(3), 1581–1591, doi: `10.1121/1.4740500`.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hearing Research **47**(1-2), 103–138, doi: `10.1016/0378-5955(90)90170-T`.

Goli, P., and van de Par, S. (**2023**). "Deep Learning-Based Speech Specific Source Localization by Using Binaural and Monaural Microphone Arrays in Hearing Aids," IEEE/ACM Transactions on Audio, Speech, and Language Processing **31**, 1652–1666, doi: `10.1109/TASLP.2023.3268734`.

Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differencesa)," The Journal of the Acoustical Society of America **63**(2), 511–523, doi: `10.1121/1.381751`.

Grantham, D. W., and Wightman, F. L. (**1979**). “Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation,” **65**, 1509–1517, doi: `10.1121/1.382915`.

Grimault, N., Parizet, E., Corneyllie, A., Brocolini, L., Weyn, R., and Garcia, S. (**2016**). “Real-time simulation of hearing impairment: Application to speech-in noise intelligibility,” The Journal of the Acoustical Society of America **140**(4_Supplement), 3438, doi: `10.1121/1.4971080`.

Grothe, B., Pecka, M., and McAlpine, D. (**2010**). “Mechanisms of Sound Localization in Mammals,” Physiological Reviews **90**(3), 983–1012, doi: `10.1152/physrev.00026.2009`.

Hall, J. W., Tyler, R. S., and Fernandes, M. A. (**1983**). “Monaural and binaural auditory frequency resolution measured using bandlimited noise and notched-noise masking,” The Journal of the Acoustical Society of America **73**(3), 894–898, doi: `10.1121/1.389013`.

Hauth, C. F., and Brand, T. (**2018**). “Modeling Sluggishness in Binaural Unmasking of Speech for Maskers With Time-Varying Interaural Phase Differences,” Trends in Hearing **22**, 233121651775354, doi: `10.1177/2331216517753547`.

Herrmann, S., and Dietz, M. (**2021**). “Model-based selection of most informative diagnostic tests and test parameters,” Acta Acustica **5**, 51, doi: `10.1051/aacus/2021043`.

Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). “Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments,” The Journal of the Acoustical Society of America **104**(4), 2412–2425, doi: `10.1121/1.423773`.

Hsieh, I.-H., Liu, J.-W., and Liang, Z.-J. (**2018**). “Spectrotemporal window of binaural integration in auditory object formation,” Hearing Research **370**, 155–167, doi: `10.1016/j.heares.2018.10.013`.

Jeffress, L. A. (**1948**). “A place theory of sound localization.,” Journal of Comparative and Physiological Psychology **41**(1), 35–39, doi: `10.1037/h0061495`.

Kates, J. M., and Prabhu, S. (**2018**). “The dynamic gammawarp auditory filterbank,” The Journal of the Acoustical Society of America **143**(3), 1603–1612, doi: `10.1121/1.5027827`.

Klug, J., and Dietz, M. (**2022**). “Frequency dependence of sensitivity to interaural phase differences in pure tones,” The Journal of the Acoustical Society of America **152**(6), 3130–3141, doi: `10.1121/10.0015246`.

Kolarik, A. J., and Culling, J. F. (**2010**). “Measurement of the binaural auditory filter using a detection task,” The Journal of the Acoustical Society of America **127**(5), 3009–3017, doi: `10.1121/1.3365314`.

Kollmeier, B., and Gilkey, R. H. (**1990**). “Binaural forward and backward masking: Evidence for sluggishness in binaural detection,” The Journal of the Acoustical Society of America **87**(4), 1709–1719, doi: `10.1121/1.399419`.

Kong, Q., Wang, Y., Song, X., Cao, Y., Wang, W., and Plumbley, M. D. (**2020**). “Source Separation with Weakly Labelled Data: An Approach to Computational Auditory Scene Analysis,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 101–105, doi: `10.1109/ICASSP40776.2020.9053396`.

Langford, T. L., and Jeffress, L. A. (**1964**). “Effect of Noise Crosscorrelation on Binaural Signal Detection,” The Journal of the Acoustical Society of America **36**(8), 1455–1458, doi: `10.1121/1.1919224`.

Majdak, P., Baumgartner, R., and Jenny, C. (**2020**). “Formation of Three-Dimensional Auditory Space,” in *The Technology of Binaural Understanding*, edited by J. Blauert and J. Braasch (Springer International Publishing, Cham), pp. 115–149, doi: `10.1007/978-3-030-00386-9_5`.

Majdak, P., Hollomey, C., and Baumgartner, R. (**2022**). “AMT 1.x: A toolbox for reproducible research in auditory modeling,” Acta Acustica **6**, 19, doi: `10.1051/aacus/2022011`.

Marquardt, T., and Mcalpine, D. (**2007**). “A π-Limit for Coding ITDs: Implications for Binaural Models,” in *Hearing – From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer Berlin Heidelberg, Berlin, Heidelberg), pp. 407–416, doi: `10.1007/978-3-540-73009-5_44`.

Marquardt, T., and McAlpine, D. (**2009**). “Masking with interaurally “double-delayed” stimuli: The range of internal delays in the human brain,” The Journal of the Acoustical Society of America **126**(6), EL177–EL182, doi: `10.1121/1.3253689`.

May, T., de Par, v., and Kohlrausch, A. (**2013**). “Binaural Localization and Detection of Speakers in Complex Acoustic Scenes,” doi: `10.1007/978-3-642-37762-4_15`.

McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). “A neural code for low-frequency sound localization in mammals,” Nature Neuroscience **4**(4), 396–401, doi: `10.1038/86049`.

Mendoza, L., Hall, III, J. W., and Grose, J. H. (**1995**). “Within- and across-channel processes in modulation detection interference,” The Journal of the Acoustical Society of America **97**(5), 3072–3079, doi: `10.1121/1.413105`.

Middlebrooks, J. C., and Simon, J. Z. (**2017**). “Ear and Brain Mechanisms for Parsing the Auditory Scene,” in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay, **60** (Springer International Publishing, Cham), pp. 1–6, doi: `10.1007/978-3-319-51662-2_1`.

Örnolfsson, I., Dau, T., Ma, N., and May, T. (**2021**). “Exploiting Non-Negative Matrix Factorization for Binaural Sound Localization in the Presence of Directional Interference,” in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 221–225, doi: `10.1109/ICASSP39728.2021.9414233`.

Oxenham, A. J., and Dau, T. (**2001**). “Modulation detection interference: Effects of concurrent and sequential streaming,” The Journal of the Acoustical Society of America **110**(1), 402–408, doi: `10.1121/1.1373443`.

Rabiner, L. R., Laurence, C. L., and Durlach, N. I. (**1966**). “Further Results on Binaural Unmasking and the EC Model,” The Journal of the Acoustical Society of America **40**(1), 62–70, doi: `10.1121/1.1910065`.

Remme, M. W. H., Donato, R., Mikiel-Hunter, J., Ballestero, J. A., Foster, S., Rinzel, J., and McAlpine, D. (**2014**). "Subthreshold resonance properties contribute to the efficient coding of auditory spatial cues," Proceedings of the National Academy of Sciences **111**(22), E2339–E2348, doi: `10.1073/pnas.1316216111`.

Ren, E., Ornelas, G. C., and Loeliger, H.-A. (**2021**). "Real-Time Interaural Time Delay Estimation via Onset Detection," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4555–4559, doi: `10.1109/ICASSP39728.2021.9414632`.

Shinn-Cunningham, B., Best, V., and Lee, A. K. C. (**2017**). "Auditory Object Formation and Selection," in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay, Springer Handbook of Auditory Research (Springer International Publishing, Cham), pp. 7–40, doi: `10.1007/978-3-319-51662-2_2;`.

Simon, J. Z. (**2017**). "Human Auditory Neuroscience and the Cocktail Party Problem," in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay, **60** (Springer International Publishing, Cham), pp. 169–197, doi: `10.1007/978-3-319-51662-2_7`.

Siveke, I., Ewert, S. D., Grothe, B., and Wiegrebe, L. (**2008**). "Psychophysical and Physiological Evidence for Fast Binaural Processing," Journal of Neuroscience **28**(9), 2043–2052, doi: `10.1523/JNEUROSCI.4488-07.2008`.

Sondhi, M. M., and Guttman, N. (**1966**). "Width of the Spectrum Effective in the Binaural Release of Masking," The Journal of the Acoustical Society of America **40**(3), 600–606, doi: `10.1121/1.1910124`.

Sutojo, S., May, T., and van de Par, S. (**2022**). "Segmentation of Multitalker Mixtures Based on Local Feature Contrasts and Auditory Glimpses," IEEE/ACM Transactions on Audio, Speech, and Language Processing **30**, 1249–1262, doi: `10.1109/TASLP.2022.3155285`.

Sutojo, S., Thiemann, J., Kohlrausch, A., and Van De Par, S. (**2020**). "Auditory Gestalt Rules and Their Application," in *The Technology of Binaural Understanding*, edited by J. Blauert and J. Braasch (Springer International Publishing, Cham), pp. 33–59, doi: `10.1007/978-3-030-00386-9_2`.

van de Par, S., and Kohlrausch, A. (**1999**). "Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters," The Journal of the Acoustical Society of America **106**(4), 1940–1947, doi: `10.1121/1.427942`.

van der Heijden, M., and Trahiotis, C. (**1999**). "Masking with interaurally delayed stimuli: The use of "internal" delays in binaural detection," The Journal of the Acoustical Society of America **105**(1), 388–399, doi: `10.1121/1.424628`.

Viemeister, N. F., and Wakefield, G. H. (**1989**). "Multiple looks and temporal integration," The Journal of the Acoustical Society of America **86**(S1), S23, doi: `10.1121/1.2027422`.

Wertheimer, M. (**1923**). "Untersuchungen zur Lehre von der Gestalt: II.," Psychologische Forschung **4**, 301–350.

Wu, Y., Ayyalasomayajula, R., Bianco, M. J., Bharadia, D., and Gerstoft, P. (**2021**). "SSLIDE: Sound Source Localization for Indoors Based on Deep Learning," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4680–4684, doi: `10.1109/ICASSP39728.2021.9415109`.

Yost, W. A. (**1974**). "Discriminations of interaural phase differences," The Journal of the Acoustical Society of America **55**(6), 1299–1303, doi: `10.1121/1.1914701`.

Yost, W. A., and Sheft, S. (**1989**). "Across-critical-band processing of amplitude-modulated tones," The Journal of the Acoustical Society of America **85**(2), 848–857, doi: `10.1121/1.397556`.

Zahorik, P. (**2021**). "Spatial Hearing in Rooms and Effects of Reverberation," in *Binaural Hearing: With 93 Illustrations*, edited by R. Y. Litovsky, M. J. Goupell, R. R. Fay, and A. N. Popper, Springer Handbook of Auditory Research (Springer International Publishing, Cham), pp. 243–280, doi: `10.1007/978-3-030-57100-9_9`.

CHAPTER 7

Glossary

Acronyms and Abbreviations

AFC	*psychoacoustic context:* alternative forced choice
AFC	*hearing algorithm context:* adaptive feedback cancelation
AMT	auditory modeling toolbox (Majdak *et al.*, 2022)
BAM-Q	binaural audio quality model (Flessner *et al.*, 2017)
BMLD	binaural masking level difference
CPSD	cross-power spectral density
dB	decibel
BU	binaural unmasking
DC	direct current
DOA	direction of arrival
ERB	equivalent rectangular bandwidth

ERD	equivalent rectangular duration
eMoBi-Q	efficient monaural and binaural quality model (chapter 5)
GPSM^q	generalized power spectrum model for quality (Biberger *et al.*, 2018)
HI	hearing-impaired (listeners)
HRTF	head-related transfer function
HWR	half-wave rectification
ILD	interaural level difference
ITD	interaural time difference
IPD	interaural phase difference
JND	just-noticable difference
LSO	lateral superior olive
LP	low-pass
MSO	medial superior olive
MoBi-Q	monaural and binaural quality model (Fleßner *et al.*, 2019)
MUSHRA	multiple stimulus with hidden reference and anchor
MVDR	minimum Variance Distortionless Response (beamformer)
NH	normal-hearing (listeners)
ODN	opposingly delayed noises
SDN	single-delayed noise
SOC	superior olivary complex
SNR	signal-to-noise ratio
SRM	spatial release from masking
TFS	temporal fine structure
YLD	years lived with disability

Fixed symbols

γ	complex correlation coefficient between two signals (in binaural context left and right), including
$\|\gamma\|$	coherence of the two signals
$\arg\{\gamma\}$	mean phase difference between the two signals, i.e. IPD
N_0	two-channel noise, index symbolizes the phase relation between the channel, i.e. zero degree (diotic noise)
N_π	two-channel noise, phase angle of $180° = \pi$ between the channels, i.e. antiphasic
S_π	two-channel pure tone, phase angle of $180° = \pi$ between the channels, i.e. antiphasic
ρ	interaural correlation, i.e. $\Re\{\gamma\}$
τ	time lag of correlation function and internal delay assumed by delay-line models [e.g., van der Heijden and Trahiotis (1999)]

Danksagung

Mein Dank gilt zunächst meinem Erstbetreuer Mathias Dietz für die Möglichkeit der Promotion, die interessante Themenstellung, die sehr engagierte und unterstützende Betreeung sowie die stringente, konstruktive Zusammenarbeit. Ich habe von seiner wissenschaftlichen Kompetenz über die ganze Zeit hinweg enorm profitiert und sehr viel gelernt. Auch habe ich die stets offene Bürotür immer sehr geschätzt, welche – als die Pandemie es wieder zuließ – unzählige spontane Absprachen ermöglicht hat.

Ganz besonders danke ich auch allen weiteren Kolleg*innen und Weggefährt*innen in der Arbeitsgruppe “Physiologie und Modellierung auditorischer Wahrnehmung”, insbesondere Jonas Klug, Anna Dietze, Henri Pöntynen, Rebecca Felsheim und Jörg Encke. Die gegenseitige fachliche wie menschliche Unterstützung, der schöne Gruppenzusammenhalt, die vielen lustigen Stunden – innerhalb wie außerhalb der Universität – haben die Promotionszeit zu einem nicht nur lehrreichen, sondern auch schönen Lebensabschnitt gemacht, aus welchem ich sehr viel mitnehme.

Weiterhin bedanke ich mich herzlich bei Thomas Biberger für die angenehme, unterstützende und lehrreiche Zusammenarbeit. Diese hat entscheidend dazu beigetragen, dass die Anwendungsforschung in meiner Disssertation die Grundlagenforschung so gut ergänzen konnte.

Stephan Ewert danke ich für die vielen wertvollen Anregungen zu den Modellen und den Manuskripten sowie für die vielen lehrreichen Treffen.

Ich bedanke mich zudem bei Volker Hohmann und dem Sonderforschungsbere-

ich Hörakustik für das konstruktive Miteinander und die inspirierende, gute wissenschaftliche Praxis, sowie insbesondere für die Unterstützung und das Interesse an meiner Arbeit trotz Nicht-Verlängerung des Teilprojekts.

Birger Kollmeier danke ich für das Verfassen des zweiten Gutachtens und den vielfachen interessanten Austausch.

Ein weiteres Dankeschön geht an die Graduiertenschule OLTECH für die zahlreichen, wertvollen Weiterbildungsangebote und Mark Pottek für die administrative Unterstützung. Die Beschäftigung mit vielerlei “Transferable Skills” war für mich ein wichtiges Element der Promotion.

Schließlich gilt mein ganz besonderer Dank meiner Familie und weiteren mir nahestehenden Menschen, die immer für mich da waren.