# Human auditory spectrotemporal processing in noise and its relationship to speech perception

Doctoral Thesis

**David López Ramos**

Supervisors

Enrique A. López-Poveda

Luis Enrique López Bascuas

July 2024

Instituto de Neurociencias de Castilla y León

Universidad de Salamanca

**A MIS ABUELOS**
**GLORIA, JOSÉ Y MARÍA**

# AGRADECIMIENTOS

Me gustaría empezar mostrando mi más sincero agradecimiento a mis dos directores: Enrique y Luis Enrique. Parafraseando una frase ya usada por Isaac Newton: "*Si he podido ver más allá es porque estoy subido en hombros de gigantes*". No se me ocurren mejores hombros en los que poder haberme subido para iniciar mi camino en la investigación científica. Luis Enrique, gracias por servirme de inspiración, por incentivar en mí la pasión por la ciencia y la valentía necesaria para realizar una tesis doctoral. Gracias por sentar las bases de mi conocimiento en el campo de la psicoacústica y por ser un gran apoyo durante estos años. Enrique, son innumerables las cosas que he aprendido de ti durante estos años. Gracias por transmitirme tu pasión y tu conocimiento, por servir de ejemplo y de inspiración. Gracias por tu confianza y por ayudarme a desarrollar habilidades que me parecían imposibles cuando empecé la tesis doctoral. He aprendido mucho de ti durante todo este tiempo y no puedo estar más agradecido por ello.

Gracias a mis compañeros de laboratorio, por hacerme sentir como en casa, ha sido un placer haber trabajado a vuestro lado. Almudena y Peter, gracias por vuestra disponibilidad, por vuestra ayuda y vuestros consejos. Habéis estado siempre que lo he necesitado. Mila y Miriam, sois las mejores hermanas mayores que uno podría tener, gracias por vuestra paciencia, apoyo y por acogerme tan bien desde el primer día. Fer, Sónia y Paula, ha sido un placer compartir con vosotros esta etapa, gracias por vuestro apoyo durante este camino que hemos recorrido juntos.

Este trabajo no hubiera sido posible sin la labor voluntaria de los participantes en los estudios que conforman la tesis doctoral. Gracias a todos ellos por su paciencia y su tiempo. También quiero agradecer a Wendy y Paula su ayuda con la recogida de datos.

Omar y Jesús, ha sido una experiencia muy enriquecedora compartir esta etapa con vosotros. Gracias por compartir conmigo vuestra pasión por la ciencia. Sin duda, sois una gran inspiración para mí. Gracias por vuestro apoyo, aun en la distancia, habéis estado siempre presentes.

Carmen, gracias por apoyarme en cada decisión vital que he tomado a lo largo de todos estos años. Es una suerte tenerte cerca y has hecho que esta etapa sea mucho más fácil. Has estado ahí cuando más lo necesitaba y eso no lo voy a olvidar nunca. Gracias por todo.

Por último, quiero dar las gracias a mis padres. Esta tesis es el resultado de mucho trabajo y esfuerzo, no sólo mío, sino también vuestro. Sois un ejemplo para mí y estoy muy orgulloso de vosotros. Vuestra honestidad y bondad, pero también vuestro coraje y sacrificio, son los mejores valores que me habéis podido inculcar. Esto es gracias a vosotros.

*David López Ramos, julio de 2024*

# FINANCIAL SUPPORT

Work supported by the Spanish Ministry of Science and Innovation (grant PID2019-108985GB-I00 to EALP) and the European Regional Development Fund. The author was hired on a four-year doctoral contract funded by the University of Salamanca and Banco Santander.

# ABSTRACT

The auditory system adapts to the varying acoustic environment to ensure a reliable and efficient representation of perceptually relevant sound features. Physiological mechanisms such as the medial olivocochlear reflex (MOCR), the middle-ear muscle reflex (MEMR) or neural synaptic adaptation can be activated by background noise, and may change the neural encoding of sound over time. Therefore, auditory frequency selectivity, modulation sensitivity, and speech perception may change dynamically in noisy settings. The overall aim of this thesis was to investigate potential noise-induced changes in frequency selectivity and modulation sensitivity, and their possible relationship with adaptation to noise in word recognition.

The thesis comprises three studies. The first study explored whether the stimulus duration affects estimates of human cochlear tuning at low and high frequencies. Cochlear tuning was estimated using a forward-masking, notched-noise (NN) method. Masker levels at masking threshold were measured for masker durations of 30 and 400 ms. The hypothesis was that tuning estimates may be different when the maskers are long enough to activate the MEMR and/or the MOCR. Masker duration had a negligible effect on tuning at 4 kHz. In contrast, the use of short masker produced broader tuning estimates at 500 Hz.

The second study investigated the effect of noise precursors on human psychoacoustical tuning curves (PTCs) at 500 Hz and 4 kHz. The stimuli used to measure the PTCs were too short (30-ms maskers and 10-ms probes) to activate the MEMR or the MOCR by themselves. The noise precursors (300 ms broadband noise at 60 dB SPL) were designed to activate the MOCR with minimal activation of MEMR. Contralateral precursors had a negligible effect on the PTCs. Ipsilateral and bilateral noise precursors sharpened the PTCs at 500 Hz but broadened the PTCs at 4 kHz. An existing physiologically inspired computer model of forward masking with efferent control was used to simulate and interpret the results. The simulated PTCs were broadly consistent with the experimental PTCs and suggested that, while seemingly different, the pattern of results at the two test frequencies is consistent with the precursors inhibiting the gain of the cochlear amplifier by activation of the MOCR.

The third study investigated if adaptation to noise occurs in spectral (SM), temporal (TM), and spectrotemporal modulation (STM) detection as well as in speech recognition. Vocoded-word recognition, TM and SM sensitivity in noise improved when the word or the modulated signal were delayed 800 ms from the noise onset. In contrast, this improvement did not occur in natural-word recognition or STM detection. Combined, these findings suggest that adaptation to noise in speech recognition is unlikely mediated by improvements in the encoding of STM.

Together, the findings show that human auditory frequency selectivity and the sensitivity to spectral and temporal modulations can change in the presence of noise. This has implications for both understanding human auditory frequency selectivity and designing optimal methods and stimuli to estimate it. Additionally, these results highlight the importance of incorporating noise adaptation mechanisms into models of the auditory periphery to accurately simulate auditory spectrotemporal processing in noisy environments.

**Keywords:** adaptation to noise, dynamic range adaptation, auditory frequency selectivity, cochlear tuning, medial olivocochlear reflex, middle-ear muscle reflex, speech-in-noise perception, spectrotemporal modulation perception.

# RELATED PUBLICATIONS

The work presented in this thesis has produced the following publications and conference presentations. The underlined text indicates the person who gave the conference presentation. A copy of the publications is given in Appendix 1.

## PEER-REVIEWED JOURNAL PAPERS

1. **López-Ramos, D.**, Marrufo-Pérez, M.I., Eustaquio-Martín, A., López-Bascuas, L.E. and Lopez-Poveda, E.A. (**in press**). Adaptation to noise in spectrotemporal modulation detection and word recognition. *Trends in Hearing*. Submitted 27 November 2023. Accepted 12 June 2024. https://doi.org/10.1177/23312165241266322

2. **López-Ramos, D.**, Eustaquio-Martín, A., López-Bascuas, L.E. and Lopez-Poveda, E.A. (**in press**). Effect of stimulus duration on estimates of human cochlear tuning. *Hearing Research. 451,* 109080. Submitted 20 March 2024. Accepted 10 July 2024. https://doi.org/10.1016/j.heares.2024.109080

3. **López-Ramos, D.**, López-Bascuas, L.E., Eustaquio-Martín, A. and Lopez-Poveda, E.A. (**under review**). Effects of ipsilateral, contralateral, and bilateral noise precursors on psychoacoustical tuning curves in humans. *Hearing Research.* Submitted 13 June 2024.

## POSTER PRESENTATIONS

1. **López-Ramos, D.**, López-Bascuas, L.E., Eustaquio-Martín, A. and Lopez-Poveda, E.A. Effects of the Medial Olivocochlear Reflex on Psychoacousical Tuning Curves. Jornadas Virtuales de la Sociedad Española de Psicología Experimental, April 22nd **2021**, Online.

2. **López-Ramos, D.**, López-Bascuas, L.E., Eustaquio-Martín, A. and Lopez-Poveda, E.A. Effects of the Medial Olivocochlear Reflex on Human Cochlear Tuning. 45th Annual MidWinter Meeting of the Association for Research in Otolaryngology, February 5th-9th 2022, Online.

3. **López-Ramos, D.**, López-Bascuas, L.E., Eustaquio-Martín, A. and Lopez-Poveda, E.A. Effect of Masker Duration on Behavioral Estimates of Human Cochlear Tuning. 46th Annual MidWinter Meeting of the Association for Research in Otolaryngology, February 11th-15th **2023**, Orlando, FL, USA.

4. **López-Ramos, D.**, López-Bascuas, L.E., Marrufo-Pérez, M.I., Eustaquio-Martín, A. and Lopez-Poveda, E.A. Adaptation to Noise in the Detection of Spectral, Temporal and Spectrotemporal Modulations. 46th Annual MidWinter Meeting of the Association for Research in Otolaryngology, February 11th-15th **2023**, Orlando, FL, USA.

5. **López-Ramos, D.**, López-Bascuas, L.E., Eustaquio-Martín, A. and Lopez-Poveda, E.A. Effect of Masker Duration on Behavioral Estimates of Human Cochlear Tuning. Audiological Research Cores in Europe (ARCHES) meeting, October 30th-31st **2023**, Salamanca, CyL, Spain.

## PODIUM PRESENTATIONS

1. **López-Ramos, D.**, López-Bascuas, L.E., Eustaquio-Martín, A. and Lopez-Poveda, E.A. Effects of ipsilateral, contralateral and bilateral medial olivocochlear reflex on psychoacoustical tuning curves. Audiological Research Cores in Europe (ARCHES) meeting, March 28th-29th, **2022**, Amsterdam, Netherlands.

2. **López-Ramos, D.**, López-Bascuas, L.E., Eustaquio-Martín, A., Marrufo-Pérez, M.I. and Lopez-Poveda, E.A. Adaptation to Noise in the Detection of Spectral, Temporal and Spectrotemporal Modulations. Audiological Research Cores in Europe (ARCHES) meeting, November 21st-22nd, **2022**, Oldenburg, Germany.
3. **López-Ramos, D.**, López-Bascuas, L.E., Eustaquio-Martín, A. and Lopez-Poveda, E.A. Effect of Masker Duration on Behavioral Estimates of Human Frequency Selectivity at Low Frequencies. 10th Iberian Conference on Perception, June 26th-28th **2024**, Madrid, Spain.

# ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| AM: | amplitude modulated/amplitude modulation. |
| AN: | auditory nerve. |
| ANOVA: | analysis of variance. |
| BM: | basilar membrane. |
| $BW_{10dB}$: | 10-dB-down bandwidth. |
| $BW_{3dB}$: | 3-dB-down bandwidth. |
| CAS: | contralateral acoustic stimulation. |
| CEOAE: | click-evoked otoacoustic emission. |
| CF: | characteristic frequency. |
| c/o: | cycles per octave. |
| dB: | decibel. |
| DRA: | dynamic range adaptation. |
| EDRNL: | dual-resonance nonlinear filter with efferent control. |
| ERB: | equivalent rectangular bandwidth. |
| FM: | frequency modulation. |
| FS: | full scale. |
| HI: | hearing impaired. |
| HL: | hearing level. |
| HSD: | honestly significant difference. |
| Hz: | Hertz. |
| IHC: | inner hair cell. |
| KEMAR: | Knowles Electronics manikin for acoustics research. |
| MEMR: | middle-ear muscle reflex. |
| MOCR: | medial olivocochlear reflex. |
| ms: | millisecond. |
| NH: | normal hearing. |
| NN: | notched noise. |
| NNTC: | notched-noise tuning curve. |
| OHC: | outer hair cell. |
| PTC: | psychoacoustical tuning curve. |
| RMANOVA: | repeated measures analysis of variance. |
| RMSE: | root-mean-square error. |
| Roex: | rounded exponential. |
| SD: | standard deviation. |
| SFOAE: | stimulus-frequency otoacoustic emission. |
| SL: | sensation level. |
| SM: | spectral modulation. |
| SNR: | signal-to-noise ratio. |
| SPL: | sound pressure level. |
| SRT: | speech reception threshold. |
| STM: | spectrotemporal modulation. |
| TFS: | temporal fine structure. |
| TM: | temporal modulation. |

# CONTENTS

**1. General introduction ..... 1**

1.1. Motivation ..... 1

1.2. Background ..... 2

1.2.1. Adaptation to noise ..... 2

1.2.2. Frequency selectivity ..... 4

1.2.3. Spectral and temporal modulation perception ..... 7

1.2.4. Speech perception in noise ..... 9

1.3. Aims and hypotheses ..... 10

1.4. Original contributions ..... 11

1.5. Overview of the thesis ..... 11

1.6. Units ..... 11

1.7. Ethics ..... 12

**2. Effect of masker duration on behavioral estimates of human cochlear tuning ..... 13**

2.1. Introduction ..... 13

2.2. Material and methods ..... 14

2.2.1. Participants ..... 14

2.2.2. Stimuli ..... 15

2.2.3. Procedure ..... 15

2.2.4. Sessions ..... 16

2.2.5. Equipment ..... 16

2.2.6. Inferring filter shape and bandwidth from notched-noise curves ..... 16

2.2.7. Statistical analysis ..... 17

2.3. Results ..... 17

2.4. Discussion ..... 21

2.4.1. Comparisons with previous human cochlear tuning estimates ..... 22

2.4.2. Factors that may contribute to the masker-duration effect at 500 Hz ..... 23

2.4.3. Factors that may contribute to the masker-duration effect at 4 kHz ..... 25

2.5. Conclusions ..... 25

**3. Effects of ipsilateral, contralateral and bilateral noise precursors on human psychoacoustical tuning curves ..... 27**

3.1. Introduction ..... 27

3.2. Material and methods ..... 29

3.2.1. Participants ..... 29

3.2.2. Stimuli ..... 29

3.2.3. Procedure ..... 30

3.2.4. Test sessions ..... 30

3.2.5. Equipment ..... 31

3.2.6. Estimating PTC bandwidths ..... 31

3.2.7. Statistical analysis ..... 31

3.2.8. Computer model ..... 31

3.3. Results ..... 32

3.3.1. Computer model simulations ..... 35

3.4. Discussion ..... 38

3.4.1. Comparisons with previous studies ..... 39

3.4.2. Model simulations ..... 40

3.4.3. Final remark on PTCs bandwidth ..... 41

3.5. Conclusions ..... 41

**4. Adaptation to noise in spectrotemporal modulation detection and word recognition ..... 43**

4.1. Introduction ..... 43

4.2. Material and methods ..... 45

4.2.1. Experimental design and hypotheses ..... 45

4.2.2. Participants ..... 45

4.2.3. Test sessions ..... 46

4.2.4. Spectrotemporal modulation task ..... 46

4.2.1. Speech recognition task ..... 48

4.2.2. Equipment ..... 49

4.2.3. Statistical analysis ..... 49

4.3. Results ..... 50

4.3.1. Thresholds in quiet ..... 50

4.3.2. Adaptation to noise ..... 50

4.4. Discussion ..... 53

4.4.1. Adaptation to noise in modulation detection ..... 53

4.4.2. Less noise adaptation for natural than for vocoded words ..... 56
4.4.3. Relationship between adaptation in modulation detection and speech recognition ..... 57
4.5. Conclusions ..... 58

**5. General discussion ..... 59**
5.1. Implications ..... 60
5.2. Limitations and future work ..... 61
5.2.1. Effect of MEMR and MOCR on frequency selectivity estimates ..... 61
5.2.2. Relationship between frequency selectivity, modulation detection, and speech-in-noise recognition ..... 61
5.2.3. Use of white noise and broadband noise with flat spectrum ..... 62
5.2.4. Adaptation to noise in SM and TM detection ..... 62
5.2.5. Lack of adaptation to noise in STM detection ..... 63
5.2.6. Lack of correlation between adaptation to noise in word recognition and modulation detection ..... 64

**Conclusions ..... 67**

**References ..... 69**

**Appendix 1. Extended summary in Spanish**

# 1.

# GENERAL INTRODUCTION

## 1.1. MOTIVATION

The auditory system encodes and processes the physical properties of sound that are relevant to perception. This is done dynamically, as the auditory system adapts to the varying characteristics of the acoustic environment, to ensure a representation of the spectrotemporal features of sounds that is both reliable and efficient (see, e.g., Willmore and King, 2023). It is therefore surprising that important aspects of auditory perception such as frequency selectivity, the sensitivity to spectral (SM), temporal (TM) and spectrotemporal modulations (STM), or their relationship to speech perception are often studied as if they were static and not subject to adaptation.

Several physiological mechanisms such as the medial olivocochlear reflex (MOCR), the middle-ear muscle reflex (MEMR) or neural synaptic adaptation could change how sounds are encoded in the auditory system over time. These mechanisms can cause frequency selectivity, modulation sensitivity and speech perception to change dynamically in noisy settings. Psychoacoustical studies have shown that human frequency selectivity can change in the presence of ipsilateral (Strickland, 2001; Jennings et al., 2009; Jennings and Strickland, 2012b) and contralateral noises (Kawase et al., 2000; Quaranta et al., 2005; Vinay and Moore, 2008; Aguilar et al., 2013; Wicher, 2013; Wicher and Moore, 2014). It has also been shown that the detection of amplitude modulation (AM) for tones in noise (Almishaal et al., 2017; Marrufo-Pérez et al., 2018b; Wojtczak et al., 2019) and the recognition of speech in noise (Ben-David et al., 2012, 2016; Marrufo-Pérez et al., 2018a) can improve as the signals are delayed in the noise.

The existing research on noise adaptation is nevertheless limited and many aspects about the effects of adaptation on auditory perception remain unknown or unclear. For instance, it remains unknown whether noise adaptation affects frequency selectivity in the same way at high and low frequencies or for adapting noise ipsilateral, contralateral, or bilateral to the test ear. Also, it is unclear if or to what extent the effects of noise adaptation on frequency selectivity are consistent with any of the physiological mechanisms mentioned earlier. Regarding modulation sensitivity, it is yet to be investigated if the sensitivity to SM, TM, and STM improves when the modulated signals are delayed in the noise. Lastly, while it is well established that speech recognition improves as the speech is delayed in the noise, it remains unclear whether this improvement is related to an enhanced representation of the speech temporal or spectral envelope, as suggested in previous studies (Ainsworth and Meyer, 1994; Marrufo-Pérez et al., 2018a, 2020).

The overall aim of this thesis is to investigate potential noise-induced changes in frequency selectivity and modulation sensitivity, and their possible relationship with adaptation to noise in word recognition.

## 1.2. BACKGROUND

### *1.2.1. Adaptation to noise*

The auditory system constantly adapts its response to the acoustic environment (reviewed by Willmore and King, 2023). This adaptation could improve the hearing function as the sounds of interest are delayed in the background noise, a phenomenon originally termed "overshoot" (Elliot, 1965; Zwicker, 1965a, b) but more recently referred to as adaptation to noise (Cervera and Ainsworth, 2005; Cervera and Gonzalez-Alvarez, 2007; reviewed in Marrufo-Pérez and Lopez-Poveda, 2022).

Noise adaptation probably results from noise-induced shifts in the dynamic range of auditory neurons towards the background noise level (Evans, 1975) and/or towards the most common level in the environment (Dean et al., 2005). In a noise background, auditory neurons adjust their dynamic range to minimize discharge rate saturation, and thus improve the encoding of both the temporal (**Fig. 1.1A**) and the spectral envelope (**Fig. 1.1B**) of the stimuli. It is not clear whether dynamic range adaptation (DRA) towards the background noise level and DRA towards the most common level in the environment are fundamentally different or essentially the same thing. DRA towards the background sound level has been studied by measuring neural responses to short tones in continuous background noise. In this situation, auditory neurons decrease their maximum firing rate and shift their dynamic range towards the noise level. This phenomenon occurs throughout the auditory system (auditory nerve (AN): Evans, 1975; Costalupes et al., 1984; Gibson et al., 1985; cochlear nuclei: Gibson et al., 1985; May and Sachs, 1992; inferior colliculus: Rees and Palmer, 1989; auditory cortex: Phillips and Hall, 1986). In contrast, DRA towards the most frequent level in the environment has been studied by measuring rate-level functions for auditory neurons in response to tones that dynamically change in level according to a pre-set distribution of levels, in which a set of levels was presented 80% of the time, constituting a "high probability region" (Dean et al., 2005). In this situation, auditory neurons adapt their dynamic range towards the levels in the high probability region. This form of DRA has also been found along the entire auditory system (AN: Wen et al., 2009, 2012; inferior colliculus: Dean et al., 2005, 2008; auditory cortex: Watkins and Barbour, 2008, 2011).

Dynamic range adaptation takes between 100 and 400 ms to occur (Wen et al., 2012). Several physiological mechanisms with roughly this time course have been suggested to cause DRA, including the MEMR (Costalupes et al., 1984; Gibson et al., 1985; Grange et al., 2022), the MOCR (Dean et al., 2005, 2008; Wen et al., 2009, 2012; Zilany and Carney, 2010; Grange et al., 2022), and synaptic processes related to exocytosis and endocytosis of neurotransmitter vesicles between inner hair cells (IHCs) and AN fibers (Dean et al., 2005, 2008; Wen et al., 2009, 2012; Zilany and Carney, 2010).

The activation of the MEMR reduces the amplitude of stapes vibration transmitted to the oval window (Hung and Dallos, 1972). It takes between 100 and 450 ms for the MEMR to be fully activated (Borg, 1982; Qiu and Stucker, 1998) and it can be triggered by sounds with levels above around 55 dB SPL (Feeney et al.,2017). Because the MEMR can attenuate the background level and its time course is roughly like the time course of DRA, this suggests that the MEMR could be involved in DRA. However, unpublished data discussed by Gibson et al. (1985) showed that DRA also occurs after cutting the tendons of the middle-ear muscle.

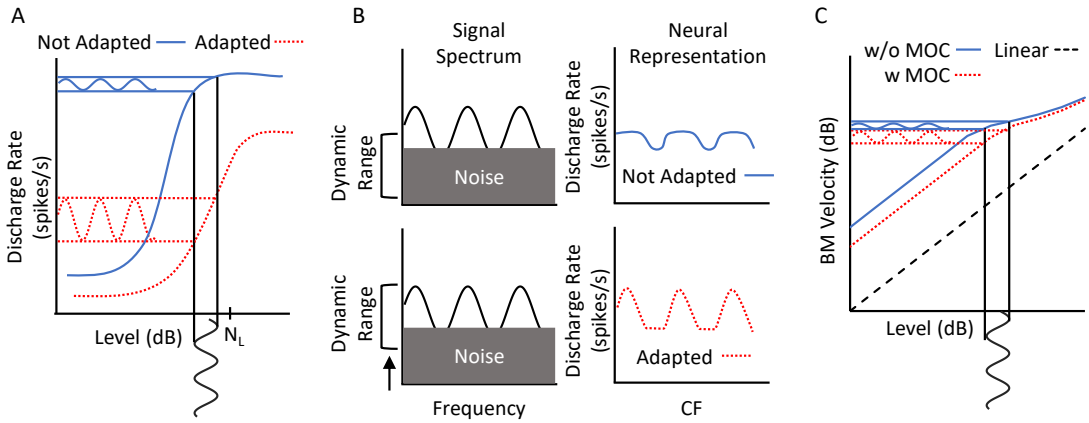


**Figure 1.1.** Schematic representation of how dynamic range adaptation (DRA) can theoretically improve the encoding of acoustic temporal and spectral modulations. The temporal envelope (black solid line) of a signal is not faithfully encoded when the neuron's response is saturated (blue solid line). However, a shift in the dynamic range towards the noise level (NL) improves the encoding of the temporal envelope (red dotted line) (**A**). Figure adapted from Marrufo-Perez et al. (2018a). Similarly, the spectral envelope (black solid line) would be poorly encoded in the neural rate profile when auditory neurons are saturated (blue solid line), but it could be more faithfully encoded thanks to DRA (red dotted line) (**B**). Figure adapted from Ainsworth and Meyer (1994). Schematic representation of basilar membrane (BM) velocity versus sound pressure level (I/O function) with and without medial olivocochlear reflex (MOCR) activation. MOCR activation linearizes BM I/O function enhancing the envelope at the output of the BM (**C**). Figure adapted from Marrufo-Perez et al. (2018a).

Medial olivocochlear neurons project from the medial regions of the superior olivary complex to the outer hair cells (OHCs) in the cochlea (Warr and Guinan, 1979; reviewed in Lopez-Poveda, 2018). The activation of medial olivocochlear neurons reduces the gain of the cochlear amplifier at low and moderate sound levels, and thus linearize basilar membrane (BM) responses (Murugasu and Russell, 1996; Dolan et al., 1997; Cooper and Guinan, 2003, 2006). In noise backgrounds, the noise could linearize the BM response by activation of the ipsilateral and/or contralateral medial olivocochlear pathways. This linearization could improve the representation of the acoustic envelope at the output of the BM (**Fig. 1.1C**). In addition, MOCR activation decompresses and shifts the rate-level function of AN fibers to higher levels (Winslow and Sachs, 1987, 1988). There is some evidence supporting the role of MOCR in adaptation to noise. For example, the MOCR presents different activation time constants, but, overall, the global time constant of MOCR activation is around 300 ms (Backus and Guinan, 2006), which is consistent with the time course of DRA for AN fibers. Additionally, when the olivocochlear bundle is cut, AN fibers do not show DRA (Kawase et al., 1993). However,

adaptation to noise in speech recognition (Marrufo-Perez et al., 2018a) and AM detection (Marrufo-Perez et al., 2019a) has been found to occur for users of cochlear implants, for whom MOCR effects are absent because the electrical stimulation delivered by the implant bypasses the OHCs (Lopez-Poveda et al., 2016).

Physiological processes at the synapses between IHCs and AN fibers, or between central auditory neurons, could also contribute to DRA. For example, the release of neurotransmitters occurs over a wide range of timescales. The fast kinetic component of neurotransmitter exocytosis at the IHC-AN synapse that underlies AN short-term adaptation is probably too fast to contribute to DRA, but the slower kinetics components, whose time course is several hundred milliseconds, are promising candidates to cause DRA (Dean et al., 2005, 2008; Wen et al., 2009, 2012; Zilany and Carney, 2010). Postsynaptic receptor desensitization could also contribute to DRA (Wen et al., 2009, 2012; Zilany and Carney, 2010).

Inhibitory processes like cochlear mechanical suppression (Cooper, 1996) or the modulation of AN fiber responses by activation of lateral olivocochlear neurons are unlikely involved in DRA because their time course is too short (almost instantaneous for suppression, Ruggero et al., 1992; Tavartkiladze et al., 1996) or too long [in the order of minutes for lateral olivocochlear neurons activation, Dean et al. (2008), Wen et al. (2009) and Zilany and Carney (2010)] compared to the time course of DRA.

Adaptation to noise has been measured in various auditory tasks such as pure tone detection (also referred as overshoot) (Zwicker, 1965a, b; Elliott, 1965; Carlyon and White, 1992; Bacon and Savel, 2004; Fletcher et al., 2013), in the detection of AM superimposed on a pure tone (Almishaal et al., 2017; Marrufo-Pérez et al., 2018b; Wojtczak et al., 2019), in frequency modulation detection (Mesik and Wojtczak, 2020), and in speech recognition (Cervera and Gonzalez-Alvarez, 2007; Ben-David et al., 2012; Marrufo-Pérez et al., 2018a) (reviewed by Marrufo-Pérez and Lopez-Poveda, 2022). However, as will be reviewed in the following sections, adaptation to noise in other important areas of auditory perception, such as frequency selectivity, SM and STM detection, or the relationship between adaptation to noise in those aspects with adaptation in speech recognition remains poorly understood.

### *1.2.2. Frequency selectivity*

Auditory frequency selectivity is the ability to separately perceive the multiple frequency components of a complex sound. This ability partly reflects the frequency tuning of AN fibers (Fletcher, 1940; Evans and Wilson, 1973; Evans, 2001), which is largely determined by the tuning of cochlear mechanical responses (Narayan et al., 1998; but see Recio-Spinoso and Oghalai, 2017).

The cochlea acts like a frequency analyzer. Pure tones produce a pattern of vibration along the BM, known as the traveling wave, whose peak occurs at different places along the BM depending on the tone frequency: high-frequency tones produce peak responses near the base of the cochlea, while low-frequency tones peak near the apex (Békésy, 1960). Cochlear tuning is closely linked to the cochlear amplifier (Rhode, 1971; reviewed in Davis, 1983), a process by which OHCs amplify the vibration of the BM and sharpen its response to specific frequencies. Cochlear amplification is also linked to compression, the phenomenon by which a wide range

of acoustics pressure is mapped (compressed) into a narrower range of cochlear mechanical (and neural) responses (Ruggero and Rich, 1991).

Auditory masking is related to frequency selectivity, i.e., when a sound (the masker) masks another sound (the probe or signal) of a different frequency, the listener cannot separately perceive the two sounds. Because of this, masking is often used to investigate cochlear compression and frequency selectivity in humans (e.g., Patterson, 1976; Nelson and Freyman, 1984; Glasberg and Moore, 1990; Oxenham and Shera, 2003; Eustaquio-Martín and Lopez-Poveda, 2011; Jennings and Strickland, 2012a). Compression has been investigated using temporal masking curves (Nelson et al. 2001; Lopez-Poveda et al., 2003) and growth of masking curves (Oxenham and Plack, 1997).

The use of masking to estimate frequency selectivity has been framed into the power spectrum model of masking (Fletcher, 1940). This model assumes that the auditory system operates like a bank of filters, each tuned to a different frequency, whose bandwidths reflect the frequency resolving capacity of the auditory system (Fletcher, 1940, Glasberg and Moore, 1990). According to the power spectrum model of masking, in a masking task the probe signal is detected by using the auditory filter that provides the better signal-to-masker ratio at its output, and the masking threshold happens when the signal-to-masker ratio at the filter output just exceeds a criterion value. The filter in question is the one at or around the signal frequency (see below).

Two popular masking techniques to estimate frequency selectivity are psychoacoustical tuning curves (PTCs) and notched-noise tuning curves (NNTCs). A PTC is a plot of the threshold masker level (masker threshold) required to just mask a fixed-level pure-tone signal (the probe) as a function of masker frequency (Small, 1959). The masker can be either a pure tone or a narrow-band noise. For PTCs measured in simultaneous masking, the use of pure-tone maskers can lead to the detection of the probe through the phenomenon of beats (Kluk and Moore, 2004), and so the use of narrow-band maskers is recommended in that case. When the masker and the probe are presented simultaneously, the response to the probe tone can be suppressed by the masker (Sachs and Kiang 1968; Cooper, 1996). Because of this, the use of forward maskers is recommended to achieve a more reasonable estimate of frequency selectivity that reflects the underlying tuning of BM responses without the confounding effect of suppression (Shera et al., 2002).

Sometimes the filter that gives the highest signal-to-masker ratio is not the one centered at the signal frequency. When that happens, the signal could be detected via adjacent filters with better signal-to-masker ratios, a phenomenon known as off-frequency listening. To prevent off-frequency listening from affecting the estimates of frequency selectivity, the notched-noise (NN) method is employed (Patterson, 1976). The NN method minimizes off-frequency listening by using two band-limited noise maskers, one above and the other one below the signal frequency. This forces the listener to rely on the frequency region around the signal frequency, providing a more accurate estimate of the auditory filter tuned at the signal frequency. Frequency selectivity is inferred from masker thresholds for a series of spectral notches with different widths between these high- and low-frequency noise bands. A NNTC is the graphical representation of masker threshold as a function of spectral notch width. NNTCs are usually fitted assuming rounded-exponential (roex) filter shapes, and the bandwidth (e.g., the

equivalent rectangular bandwidth or ERB) of the fitted roex filter is taken as an estimate of frequency selectivity.

Using the NN method, previous studies have investigated how human frequency selectivity changes as a function of frequency (Glasberg and Moore, 1990; Oxenham and Shera, 2003; Leschke et al., 2022) (**Fig. 1.2**). The estimates of Glasberg and Moore (1990) were obtained using the NN method applied in simultaneous masking, with fixed-level maskers, and varying the probe level to estimate the masked probe threshold; that is, using an iso-stimulus approach (Eustaquio-Martín and Lopez-Poveda, 2011). This approach differs from physiological approaches where cochlear frequency selectivity is measured by varying the stimulus level to achieve a fixed response (iso-response approach). To make behavioral estimates of frequency selectivity as comparable as possible to physiological measures of cochlear tuning, Shera et al. (2002) applied the NN method in forward masking, with fixed-level probes, and varying the masker level to measure masker threshold. The estimates obtained with this latter method were narrower than those obtained by Glasberg and Moore (1990) and suggested that cochlear tuning was sharper in humans than in other mammals (see also Shera et al. (2010), for otoacoustic emissions evidence). Recently, the results from Oxenham and Shera (2003) have been extended to frequencies below 1 kHz (Leschke et al., 2022).

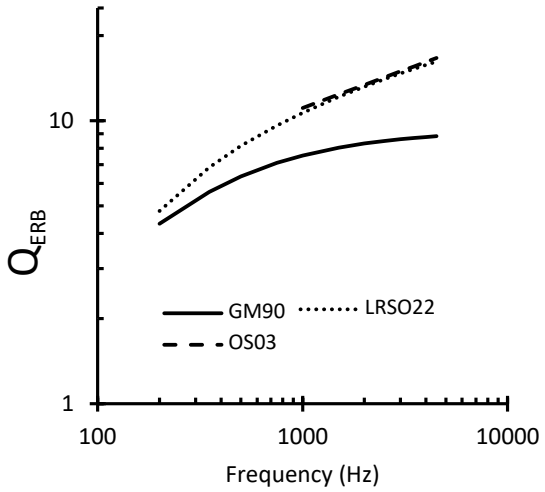


**Figure 1.2.** $Q_{ERB}$ estimates as a function of frequency. $Q_{ERB}$ = Probe frequency/ERB. Curves show regression equations reported in previous studies. GM90 (solid line): Glasberg and Moore (1990); OS03 (dashed line): Oxenham and Shera (2003); LRSO22 (dotted line): Leschke et al. (2022).

The studies just described, and the most recent behavioral estimates of human frequency selectivity were obtained disregarding the potential effects of adaptation described in **Section 1.2.1**. Adaptation could alter frequency selectivity in noisy environments. Previous studies have suggested that human frequency selectivity could be affected by contralateral acoustic stimulation (CAS) (Kawase et al., 2000; Quaranta et al., 2005; Vinay and Moore, 2008; Aguilar et al., 2013; Wicher, 2013; Wicher and Moore, 2014) or by ipsilateral sound precursors (Strickland, 2001; Jennings et al., 2009; Jennings and Strickland, 2012b). In addition, much psychoacoustical evidence [reviewed in Lopez-Poveda (2018) and Jennings (2021)] suggests that the use of sound precursors can inhibit the gain of the cochlear amplifier and linearize BM

responses as inferred using behavioral masking methods (Krull and Strickland, 2008; Jennings et al., 2009; Roverud and Strickland, 2010; Yasin et al., 2014; DeRoy Milvae and Strickland, 2018,2021; Salloom and Strickland, 2021; Salloom et al., 2023).

Although this evidence suggests that frequency selectivity may vary dynamically in noisy environments, the issue has not been systematically investigated. Furthermore, methodological differences across studies complicate the analysis of important aspects, leaving several crucial questions unanswered. For instance, the maskers used to measure frequency selectivity are often long enough and their level is high enough that they could trigger some of physiological mechanisms involved in adaptation by themselves. However, it is unclear if this is the case or to what extent behavioral estimates of frequency selectivity vary depending on masker duration. Also, it is unclear if frequency selectivity in noise is different from frequency selectivity in quiet, if it depends on the laterality of the noise, or whether the potential impact of noise on frequency selectivity is different at high and low frequencies. Also, it remains unclear if the potential effects of noise on frequency selectivity are consistent with the previously discussed adaptive processes such as the MOCR or MEMR. One aim of the present thesis is to investigate some of these issues.

### *1.2.3. Spectral and temporal modulation perception*

Another important aspect of auditory perception is the sensitivity to SM, TM, and STM. SMs pertain to variations in energy over frequency, while TMs manifest as changes in amplitude over time. STMs combine simultaneous spectral and temporal modulations, thus appear in a spectrogram as diagonal energy shift (**Fig. 1.3**).

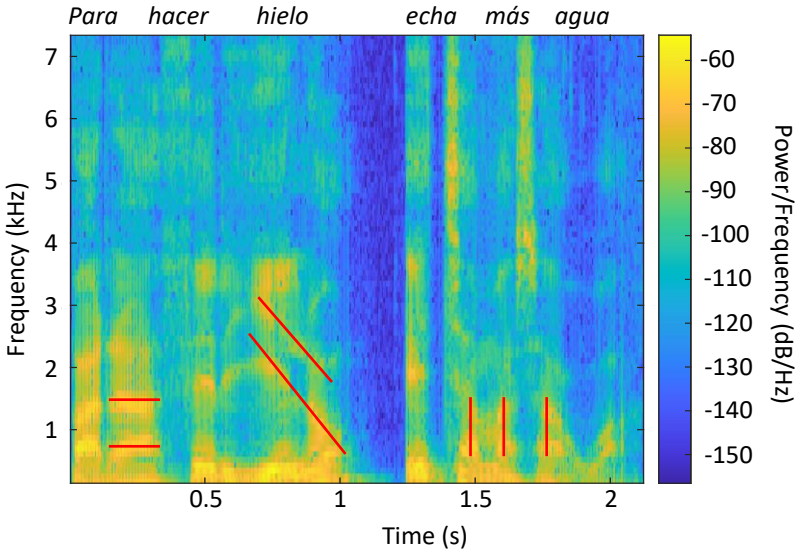


**Figure 1.3.** Spectrogram of the Spanish sentence "*Para hacer hielo echa más agua*". Red lines superimposed on the spectrogram denote spectral (SM) (horizontal), temporal (TM) (vertical), and spectrotemporal (STM) (diagonal) modulations.

There are various methods to evaluate SM, TM, and STM sensitivity. These methods typically involve using noise or a multi-tone complex with modulated spectral and temporal envelopes.

The modulation of the spectral and temporal envelopes can be characterized by the spectral density (in cycles per octave, c/o) and temporal modulation rate (in Hz), respectively, along with the phase of each type of modulation. In a STM *detection* task, listeners must differentiate a noise with modulated spectral and/or temporal envelopes from a noise with flat spectral and temporal envelopes. In this task, the dependent variable is the modulation depth (Bernstein et al., 2013,2016; Mehraei et al., 2014; Venn-Davies et al., 2015; Isarangura et al., 2019; Magits et al., 2019; Zaar et al., 2023). By contrast, in a STM *discrimination* task, listeners are asked to distinguish spectrotemporally modulated noise from noise that is also spectrotemporally modulated but with a different spectral density (Aronoff and Landsberger, 2013); or a phase-reversed modulation (Supin et al., 1998; Henry et al., 2005; Venn-Davies et al., 2015); or with a different STM direction (Archer-Boyd et al. 2018). In these cases, the modulation depth is typically fixed, and the dependent variable is the spectral modulation density.

Since SM and STM detection requires resolving spectral peaks and valleys, it is reasonable to think that SM and STM sensitivity is limited by the frequency-resolving capacity of the auditory system (Summers and Leek, 1994; Leek and Summers, 1996; Isarangura et al., 2019). At high densities of spectral peaks and valleys, SM detection could be limited by frequency selectivity. In addition, the broadening of peripheral auditory filters found in hearing-impaired (HI) listeners (Glasberg and Moore, 1986; Yonemoto et al., 1998) is associated with a poor internal representation of the spectral profile (Leek et al., 1987; Summers and Leek, 1994; Leek and Summers, 1996). Combining experimental and modeling data, Ponsot et al. (2021) showed that the difference between normal-hearing (NH) and HI listeners in spectrotemporal processing could be primarily explained by the broader cochlear filters in HI listeners. There is evidence for a relationship between SM and STM detection and frequency selectivity. For example, Venn-Davies et al. (2015) found significant correlations between SM detection thresholds at different spectral densities and the ERBs of auditory filters at 500 Hz and 2 kHz, which suggests that the performance in frequency selectivity and SM detection tasks could be governed by common mechanisms. Similarly, Bernstein et al. (2013) found that STM detection was strongly correlated with both low-frequency modulation (FM) sensitivity and the ERBs at 4 kHz. Also, Mehraei et al. (2014) found that when 4 c/o and 4 Hz STM was superimposed on a one-octave noise band centered at 4 kHz, HI listeners had significantly poorer detection thresholds than NH listeners, which is consistent with reduced frequency selectivity affecting STM sensitivity.

Sensitivity to SM, TM, and STM has been extensively studied, for example by calculating SM (Summers and Leek, 1994; Eddins and Bero, 2007; Isarangura et al., 2019), TM (Viemeister,1979) and STM (Chi et al., 1999) transfer functions, which relate modulation sensitivity to spectral modulation density, temporal modulation rate, or both simultaneously. TM detection superimposed on wide-band noises shows a low-pass trend with a relatively constant modulation sensitivity up to 16 Hz and then decline gradually for higher modulations (Viemeister,1979). SM sensitivity seems to peak around 1 and 4 c/o with a decrease in sensitivity for lower and higher spectral densities (Summers and Leek, 1994; Eddins and Bero, 2007; Isarangura et al., 2019). Data from Chi et al. (1999) show that the STM transfer function can be explained by regarding SM and TM sensitivity as two independent variables. However, this has been challenged by physiological evidence (e.g., Escabí and Schreiner, 2002).

The studies just described, and their conclusions were made without considering the dynamic adaptation processes that may affect modulation detection. It has been shown that for AM tones in noise, AM detection improves as the tones are in delayed in the noise (Marrufo-Pérez et al., 2018b) or preceded by noise precursors (Almishaal et al., 2017; Wojtczak et al., 2019) (reviewed in Marrufo-Pérez and Lopez-Poveda, 2022). This suggests that TM sensitivity may be enhanced by adaptation processes occurring in the auditory system. However, little is known about how the sensitivity to SM, TM, and STM is affected by background noise, or whether the noise adaptation mechanisms mentioned above can improve modulation sensitivity. Another aim of the present thesis is to investigate adaptation to noise in the detection of SM, TM and STM.

### *1.2.4. Speech perception in noise*

Noise adaptation in speech recognition has been studied in several ways. One method involves varying the speech-noise onset delay and measuring the improvement (if any) in the percentage of recognized words (Ainsworth and Meyer, 1994; Cervera and Ainsworth, 2005; Cervera and Gonzalez-Alvarez, 2007) or in the speech reception threshold (SRT) (i.e., the signal-to-noise ratio at 50%-word recognition) (Ben-David et al., 2012,2016; Marrufo-Pérez et al., 2018a). Another method involves the use of noise precursors. In this case, adaptation is calculated as the difference between speech scores observed with and without the noise precursor (Marrufo-Pérez et al., 2020). Previous research has shown that adaptation to noise can improve word recognition by up to 30% (Ainsworth and Meyer, 1994; Cervera and Ainsworth, 2005; Cervera and Gonzalez-Alvarez, 2007), and SRTs up to 5 dB (Ben-David et al., 2012, 2016; Marrufo-Pérez et al., 2018a). However, it is not yet clear what speech cues account for these improvements.

Speech recognition relies on the accurately encoding of the SMs, TMs, and STMs contained in speech (**Fig. 1.3**) (Elliot and Theunissen, 2009). Recognition worsens when speech is degraded either spectrally (Ainsworth and Millar, 1972; van Veen and Houtgast, 1985; Zahorian and Jagharghi, 1993; Liu and Eddins, 2008) or temporally (Drullman et al., 1994), however, it is impossible to reduce temporal information in speech without also impacting spectral information, and vice versa (Drullman et al., 1996). In addition, there is evidence that speech recognition worsens when low spectral (<4 cycles/kHz) and slow temporal (<12 Hz) modulations are eliminated from speech. (e.g., Elliot and Theunissen, 2009).

Several studies have investigated the relationship between speech recognition and the sensitivity to SM (Henry et al., 2005; Venn-Davies et al., 2015) and STM (Bernstein et al., 2013,2016; Mehraei et al., 2014; Zaar et al., 2023). SM detection thresholds (Venn-Davies et al., 2015) and SM discrimination thresholds (Henry et al., 2005; Venn-Davies et al., 2015) are significant predictors of speech recognition in noise. The relationship between STM and speech intelligibility depends on the spectral density, the temporal modulation rate, and the speech intelligibility measure. Overall, however, the sensitivity to STM accounts for a significant part of the variance of speech intelligibility not accounted for by measures of audibility such as the audiogram or the speech intelligibility index (Bernstein et al., 2013;2016; Mehraei et al., 2014; Zaar et al., 2023). This suggests that the improvement in word recognition reported in previous

noise adaptation studies may be due to an improved encoding of the speech spectral and/or temporal envelopes in the auditory system.

Some studies have provided evidence in support of this hypothesis. For example, the fact that AM detection for tones in noise improves over time suggests that adaptation to noise in speech recognition occurs because of an improvement in the encoding of the temporal envelope of speech. This is further supported by the fact that noise adaptation is greater for words processed to maintain only temporal envelopes cues (i.e., vocoded words) than for unprocessed words that keep envelope and temporal fine structure (TFS) cues (Marrufo-Pérez et al., 2018a; 2020). A computational modelling study (Ainsworth and Meyer, 1994) showed that when a threshold shift towards the noise level is introduced in the rate-level function of the model AN fibers, the speech spectral envelope is better encoded in the AN. When the output of the AN model was sent to an automatic speech recognizer, automatic speech recognition increased because of this improvement in the encoding of the speech spectral envelope. To date, however, no studies have investigated whether adaptation to noise in word recognition is related with an improved encoding of speech spectral or temporal cues. Another aim of the present thesis is to investigate this possibility.

## 1.3. AIMS AND HYPOTHESES

The main aim of this thesis is to experimentally investigate the effect of background noise on human auditory spectrotemporal processing and whether noise-induced effects on spectrotemporal processing are related to adaptation to noise in speech recognition. The overall hypothesis is that frequency selectivity and modulation sensitivity change in background noise due to physiological adaptation mechanisms. This change in spectrotemporal processing could be related to adaptation to noise in word recognition.

The specific objectives and hypothesis are as follows:

1. To investigate whether behavioral estimates of human cochlear tuning obtained with the most accurate NN method (Shera et al., 2002) may be affected by noise adaptation. It is hypothesized that some of maskers used with this technique are long enough and their level is high enough that they could trigger the MOCR and MEMR, which would affect the tuning estimates.
2. To investigate the potential effect of contralateral, ipsilateral, and bilateral noise precursors on frequency selectivity estimates. It is hypothesized that noise precursors could affect auditory filter bandwidth in a way consistent with an activation of the MEMR and/or MOCR by the precursors. It is also explored if the noise precursors affect frequency selectivity differently at low and high frequencies.
3. To investigate if the sensitivity to SM, TM, and STM in noise improves as the signals are delayed from the noise onset. As explained above (**Fig. 1.1**), neural DRA can improve the encoding of spectrally and temporally modulated envelopes. Therefore, it is hypothesized that the sensitivity to SM, TM and STM should improve when the modulated stimulus is delayed in the noise.
4. To investigate if and to what extent adaptation to noise in word recognition relates to an improvement in SM, TM, and/or STM sensitivity. It is hypothesized that the same listeners

who show noise adaptation in word recognition will also show adaptation in the detection of all three types of modulation.

## 1.4. ORIGINAL CONTRIBUTIONS

The main original contributions of this thesis are:

1. Empirical evidence that behavioral estimates of human cochlear tuning depend on stimulus duration at 500 Hz, but not at 4 kHz.
2. Experimental evidence that noise precursors affect behavioral estimates of frequency selectivity differently depending on the laterality of the noise and the signal frequency. In addition, computational modelling evidence that these experimental results can be explained by activation of the MOCR by the noise precursors.
3. Empirical evidence that the sensitivity to TM superimposed on a multi-tone complex improves when the carrier is delayed in noise.
4. Empirical evidence that the sensitivity to SM but not to STM improves when the modulated carrier is delayed in noise.
5. Experimental evidence that for NH listeners adaptation to noise in speech perception is not related to an improvement in sustained STM detection in noise.

## 1.5. OVERVIEW OF THE THESIS

This thesis is organized according to the objectives listed in **Section 1.3**.

**Chapter 2** describes an experimental study aimed at investigating the effect of masker duration on behavioral estimates of human cochlear tuning at 500 Hz and 4 kHz obtained using the forward-masking, NN method.

**Chapter 3** presents a study aimed at investigating the effect of ipsilateral, contralateral, and bilateral noise precursor on 500-Hz and 4-kHz PTCs measured using short-duration maskers. In addition, a physiologically inspired phenomenological computer model of forward masking with efferent control is used to simulate the effect of the precursors on the PTCs assuming that the precursors reduce the gain of the cochlear amplifier by activation of the MOCR.

**Chapter 4** describes a study aimed at investigating whether adaptation to noise occurs in the detection of SM, TM, and STM. For the same participants (repeated measured design), adaptation to noise is also measured for word recognition to investigate whether adaptation in speech recognition is related to improvements in the encoding of SM, TM and/or STM.

**Chapter 5** provides a comprehensive and unified discussion of the main findings, highlighting their significance, and discussing their implications and limitations. Furthermore, this chapter suggests new challenges and outlines potential future work.

## 1.6. UNITS

All units follow the conventions of International System of Units (SI).

## 1.7. ETHICS

The work reported in this thesis involved conducting behavioral experiments in human subjects and collecting personal data. The procedures were approved by the Ethics Review Board of the University of Salamanca (Salamanca, Spain). All participants were adult volunteers and not paid for their time. They all signed an informed consent before they were admitted to the experiments.

# 2.

# EFFECT OF MASKER DURATION ON BEHAVIORAL ESTIMATES OF HUMAN COCHLEAR TUNING[1]

## 2.1. INTRODUCTION

Auditory frequency selectivity is thought to be largely determined at the level of the auditory nerve (AN) (Fletcher, 1940; Evans and Wilson, 1973; Evans, 2001), and AN tuning is itself largely determined by the tuning of cochlear mechanical responses (Narayan et al., 1998; but see Recio-Spinoso and Oghalai, 2017). Based on this, auditory masking methods originally employed to assess behavioral frequency selectivity have evolved over the years to infer cochlear tuning.

Patterson (1976) proposed one of the most significant early behavioral methods to infer auditory filter shapes. The method consisted of measuring the detection threshold of a pure tone probe in the presence of a simultaneous notched-noise (NN) masker as a function of the notch width. Masking was assumed to occur when the ratio of the probe-to-noise power spectrum at the output of the auditory filter tuned at the probe frequency met a certain criterion. Key to this method was the use of noise maskers with a spectral notch centered at or around the probe frequency, which was intended to minimize off-frequency listening (O'Loughlin and Moore, 1981).

Shera et al. (2002) improved the method of Patterson (1976) in several ways to infer auditory filters closer to cochlear filters. First, they used forward rather than simultaneous maskers to minimize the potential confounding effects of cochlear suppression (Cooper, 1996) on tuning (Oxenham and Shera, 2003). Second, they used a fixed-level probe and varied the masker level to infer cochlear tuning in a way most similar to the way mechanical and neural tuning curves are measured, i.e., they used an isoresponse procedure (Eustaquio-Martin and Lopez-Poveda, 2011). Third, they used low-level probes to minimize the spread of cochlear excitation caused by the probe and thus infer tuning for a narrow cochlear region.

---

[1] The work described in this chapter is presented in the following publication: López-Ramos, D., Eustaquio-Martín, A., López-Bascuas, L.E. and Lopez-Poveda, E.A. (in press). Effect of stimulus duration on estimates of human cochlear tuning. *Hearing Research. 451,* 109080. Submitted 20 March 2024. Accepted 10 July 2024. https://doi.org/10.1016/j.heares.2024.109080

Shera et al. (2002) showed that their behavioral method provides tuning estimates broadly consistent with estimates inferred from the latency of stimulus-frequency otoacoustic emissions (SFOAEs), a more direct measure of the cochlear mechanical response. Sumner et al. (2018) showed that in ferrets the masking method of Shera et al. (2002) provide bandwidth estimates broadly consisted with AN tuning measured in the same animal. This supports the validity of the method of Shera et al. (2002) for inferring cochlear tuning (but see Ruggero and Temchin, 2005; Eustaquio-Martin and Lopez-Poveda, 2011; Lopez-Poveda and Eustaquio-Martin, 2013).

Studies are lacking on the effect of masker duration on tuning estimates inferred with this method. Cochlear tuning can vary, for example, depending on the state of activation of the medial olivocochlear reflex (MOCR) (Cooper and Guinan, 2006; reviewed in Lopez-Poveda, 2018 and Jennings, 2021). Furthermore, cochlear tuning varies with sound level (in the cochlear base: Ruggero et al., 1997; in the apex: Recio-Spinoso and Oghalai, 2017), and the effective level reaching the inner ear can vary depending on the state of activation of the middle-ear muscle reflex (MEMR). The MOCR and the MEMR can be elicited by ipsilateral broadband sounds longer than about 50-100 ms (Church and Cudahy, 1984; Backus and Guinan, 2006), and their effects increase with increasing level for levels above the activation threshold of the reflexes: ~40 dB SPL for the MOCR (Backus and Guinan, 2006) and ~55 dB SPL for the MEMR (Feeney et al., 2017). The noise maskers typically used in the method of Shera et al. are long enough (e.g., 400 ms in Shera et al., 2002 and Oxenham and Shera, 2003; or 200 ms in Leschke et al., 2022) and their overall level is high enough (e.g., from 30 to 88 dB SPL in Shera et al., 2002; or from 40 to 73 dB SPL in Leschke et al., 2022) that they could activate either or both reflexes. Therefore, the tuning estimates inferred with the method of Shera et al., (2002) using long maskers could be affected by the MOCR and/or the MEMR.

The aim of the present study was to investigate whether human cochlear tuning inferred with the method of Shera et al. (2002) depends on masker duration. Auditory filter shapes and tuning were inferred using short (30 ms) and long (400 ms) maskers. It was assumed that the 30-ms maskers were too short to activate the MOCR or the MEMR, while the 400-ms maskers were long enough to activate the two reflexes provided their level was high enough. The probe frequencies were 500 Hz and 4 kHz to investigate the effect of masker duration on tuning for apical and basal cochlear regions, respectively. It will be shown that tuning estimates at 4 kHz are not affected by masker duration. In contrast, at 500 Hz, the use of short maskers produces broader estimates of tuning.

## 2.2. MATERIAL AND METHODS

### *2.2.1. Participants*

Seven people (4 women) with normal clinical audiograms (thresholds ≤ 20 dB hearing level (HL); ANSI, 1996) and no self-reported history of hearing impairment participated in the experiments. Their ages were 25 (S1), 29 (S2), 30 (S3), 29 (S4), 24 (S5), 24 (S6), and 23 (S7) years. Participants were tested in their right ears, except S5 who was tested in her left ear because of a mild hearing loss in her right ear. Participants were volunteers and were not paid

for their time. Except S6 and S7, all other participants were experienced in psychoacoustic tasks.

### 2.2.2. Stimuli

Masker levels at the probe detection thresholds (hereinafter referred to as masker thresholds) were measured in a forward masking paradigm. Probes were sinusoids at 10 dB sensation level (SL) and with a duration of 20 ms, including 10-ms cosine-squared rise/fall ramps and no steady-state portion. The duration of the masker was either 30 ms or 400 ms, including 10-ms cosine-squared rise/fall ramps. There was no time gap between the masker and the probe. The NN maskers consisted of two bands of Gaussian noise centered below and above the probe frequency ($f_P$). The noise bands were generated in the spectral domain, i.e., obtaining their waveforms by an inverse fast Fourier Transform, which results in infinitely steep spectral slopes. The width of the bands was set to $0.25 \times f_P$. Masker thresholds were measured for different relative notch widths, $\Delta g$, defined as the spectral distance between the closest edges of each of the band and $f_P$, relative to $f_P$, i.e., $\Delta g = | f - f_P | / f_P$. Masker thresholds were measured for five symmetric (0, 0.1, 0.2, 0.3, 0.4) and two asymmetric (0.4 and 0.2; 0.2 and 0.4) relative notch widths. For some participants, masker levels for the wider notches reached the maximum output level of the equipment (100 dB SPL) when the probe frequency was 4 kHz and the masker duration was 30 ms. In those cases, masker thresholds were measured for three additional relative notch widths: 0.15, 0.25 and asymmetrical notch (0.3 and 0.1).

### 2.2.3. Procedure

A three-interval, three-alternative, forced-choice adaptive procedure with feedback was employed to estimate each masker threshold. The initial masker level was set at 20 dB SPL (a level presumably below the estimated masker threshold) and the masker level was then changed according to a two-up, one-down adaptive procedure to estimate the 71% point on the psychometric function (Levitt, 1971). The masker level was changed in 6-dB steps for the first two reversals and in 2-dB steps thereafter. The adaptive procedure continued until 10 reversals in masker level were recorded. The mean masker level at the last eight reversals was regarded as the threshold. The threshold was discarded if the standard deviation (SD) of the eight reversals exceeded 6 dB. Three thresholds were measured per condition. Like Leschke et al. (2022) here, a threshold was discarded if it was more than four SDs below the mean of the other two thresholds. This was intended to avoid including masker thresholds in which participants could have lost concentration, yielding a masker threshold that was far below the other thresholds. In those cases (81 of a total of 591 thresholds), the mean of the remaining two values was taken as the masker threshold.

The absolute hearing threshold for the probes were estimated using a two-down, one-up adaptive rule. Three measurements were made, and the mean was taken as the threshold. Individual absolute thresholds for the 500-Hz probe were 20 (S1), 20 (S2), 18 (S3), 20 (S4), 23 (S5), 36 (S6), and 26 (S7) dB SPL. Absolute thresholds for the 4-kHz probe were 18 (S1), 21 (S2), 13 (S3), 30 (S4), 14 (S5), 30 (S6), and 13 (S7) dB SPL.

### 2.2.4. Sessions

The first session was devoted to calculating the absolute hearing thresholds for 20-ms probe tones. The following sessions were devoted to measuring NN curves for the 500-Hz and 4-kHz probes for short (30 ms) and long (400 ms) maskers. In each session, two NN curves were measured, one per masker duration. Notched-noise curves for long and short maskers were measured in balanced order across sessions. Notched-noise curves for the 500-Hz and 4-kHz probes were also measured in a balanced order across sessions. The entire experiment took about 10 hours per participant, divided into one-hour sessions. Breaks were given within each session as demanded by the participants.

### 2.2.5. Equipment

Stimuli were generated digitally using custom-made MATLAB software (R2017b, The Mathworks) and played via an RME Fireface UCX soundcard at a sampling rate of 44.1 kHz and 24-bit resolution. Stimuli were presented through Etymotic Research ER2 insert earphones designed to give a flat frequency response at the eardrum between 125 Hz and 10 kHz. The equipment was calibrated at 1 kHz by placing the earphones on a KEMAR (Knowles Electronics) equipped with a Zwislocki DB-100 artificial ear connected to a sound level meter (Brüel Kjaer, mod. 2238), and the obtained sensitivity was used at all other frequencies. During the experiments, participants were seated in a double-walled sound-attenuating chamber.

### 2.2.6. Inferring filter shape and bandwidth from notched-noise curves

Notched-noise curves were fitted assuming rounded-exponential (roex) filter shapes and the power spectrum model of masking (Patterson, 1976). Methods were identical to those employed in earlier studies (Shera et al., 2002; Oxenham and Shera, 2003; Leschke et al., 2022) except that here the human middle ear filter was based on data from Goode et al. (1994) (Fig. 1, set: 104 dB SPL). Because the insert earphones used here (Etymotic Research ER2) are designed to produce an approximately flat response at the eardrum, no outer ear filter was applied.

The bandwidths were estimated from the individual and mean NNTCs by assuming two different roex: roex(*pwt*) and roex(*pwtp*). The first one is defined by **Eq. (2.1)**:

$$W(g) = (1 - w)(1 + pg)e^{-pg} + w\left(1 + \frac{pg}{t}\right)e^{-\frac{pg}{t}} \quad (2.1)$$

where $g$ is the frequency relative to the probe frequency ($f/f_P$). The roex(*pwt*) has two slopes, one at the tip, $p$, and one at the tail, $t$. The intersection of these two slopes is set by the parameter $w$ ($0 \leq w \leq 1$). The value of $p$ was allowed to differ for the lower and upper sides of the filter, while $t$ and $w$ were assumed to have the same value on both sides of the filter.

The roex(*pwtp*) was identical to the roex(*pwt*) on the lower side, but its upper side was assumed to have only a single slope define by the parameter $p_u$, as shown by **Eq. (2.2)**:

$$W(g) = (1 + p_u g)e^{-p_u g} \quad (2.2)$$

The parameters of the roex functions were optimized as described in previous studies (Oxenham and Shera, 2003; Eustaquio-Martín and Lopez-Poveda, 2011; Lopez-Poveda and Eustaquio-Martín, 2013). It was assumed that the probe was detected with the filter centered at the signal frequency. Because the roex fit to the NN curve (and hence, the bandwidth inferred from the fit) may depend on the initial values of the parameters in the optimization process, the fitting procedure for each roex type to each NN curve was repeated several times with random initial parameters until we found 100 fits with root-mean-squared error (RMSE) ≤ 6 dB. Of the resulting 100 models, the one with the best fit (smallest RMSE) was chosen. Following the recommendations of previous studies (e.g., Glasberg and Moore, 1990; Oxenham and Shera, 2003), models with extreme values of tip slopes ($p>200$) or having negative $t$ values were discarded even though some of them produced better fits than the ones reported here.

For each condition and participant, the equivalent rectangular bandwidth (ERB) was calculated for by integration of the roex filter shape, and the corresponding filter quality factor, $Q_{ERB}$, was calculated as $Q_{ERB}= f_P /ERB$.

### *2.2.7. Statistical analysis*

Statistical analyses were performed using R (version 4.1.0). The Shapiro-Wilk test of normality revealed that the ERBs of the 4-kHz auditory filters were normally distributed. Therefore, a one-way repeated measures ANOVA was applied to assess the statistical significance of the effect of masker duration on the ERB at 4-kHz. ERB values for the 500-Hz auditory filters did not follow a Gaussian distribution. Therefore, a Wilcoxon signed rank test was applied. We also checked whether the ERB and the fit error (RMSE) differed for the two roex [i.e., roex($pwt$) vs roex($pwtp$)]. At 4 kHz, the ERBs followed a normal distribution but the RMSEs did not. Therefore, a Student $t$ test and a Wilcoxon signed rank test were used to analyze the effect of roex type on the ERB and RMSE, respectively. At 500 Hz, Wilcoxon signed rank tests were used to analyze the effect of roex type on the two variables (ERB and RMSE) because neither of them followed Gaussian distributions at this frequency.

## 2.3. RESULTS

**Figure 2.1** and **Fig. 2.2** illustrate the NN curves for probe frequencies of 500 Hz and 4 kHz, respectively. Circles and triangles depict masker thresholds for symmetric and asymmetric notches, respectively. For the 500-Hz probe (**Fig. 2.1**), NN curves are overall shallower for the 30-ms than for the 400-ms masker. Furthermore, masker thresholds for asymmetric notches were coincident for the two masker durations (left- and right-pointing triangles), which suggest roughly symmetric filters for the two masker durations. For the 4-kHz probe (**Fig. 2.2**), NN curves for the 30-ms masker are shifted upwards relative to those for the 400-ms masker. For most participants, thresholds were different for the two asymmetric notch conditions for the 400-ms masker. For the 30-ms masker, in one of the asymmetric notch conditions (lower edge: $0.4f_P$ upper edge: $0.2f_P$), the masker level reached the maximum output of the equipment for some listeners, while for the other asymmetric notch condition (lower edge: $0.2f_P$ upper edge:

0.4$f_P$) thresholds never reached this limit. These results suggest that the asymmetry of auditory filters at 4 kHz is different for the two masker durations.

The two types of roex functions used to infer filter shapes from the NN curves did not differ in ERB (at 500 Hz: Wilcoxon test Z=44, p=0.626; at 4 kHz: paired Student's *t* test t(13)=-1.88, p=0.083) nor RMSE (at 500 Hz: Wilcoxon test Z=33, p=0.666; at 4 kHz: Wilcoxon test Z=39, p=0.426). For this reason, in what follows, only roex(*pwtp*) estimates are shown. The roex(*pwtp*) parameters and the RMSEs for each participant and masker duration are shown in **Table 2.1** for 500-Hz probe and in **Table 2.2** for 4-kHz probe.

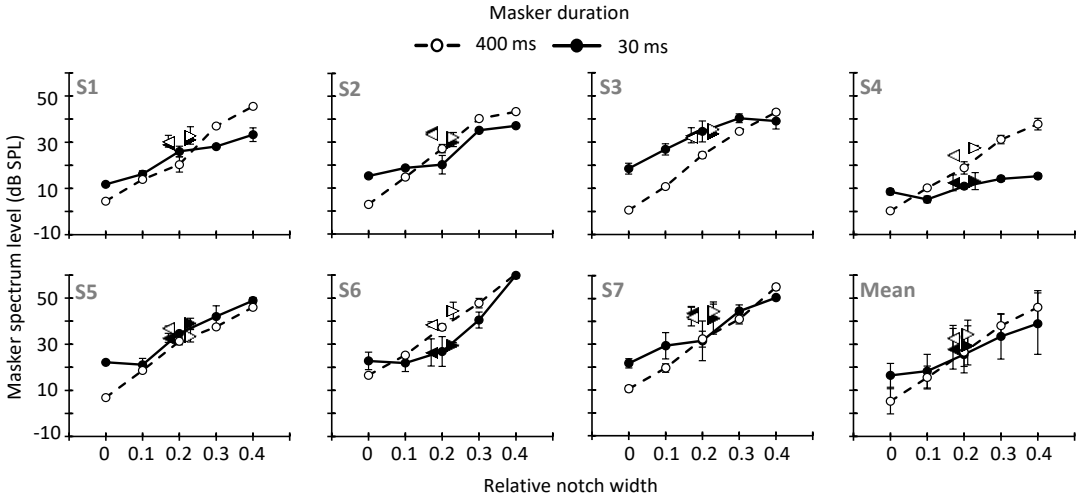


**Figure 2.1.** Individual and mean notched-noise tuning curves (NNTCs) for a probe frequency of 500 Hz. Each panel illustrates results for an individual participant or the mean, as indicated in the top-left corner of each panel. Filled symbols and solid lines depict masker thresholds for the short, 30-ms masker; open symbols and dashed lines depict thresholds for the long, 400-ms masker. Circles and triangles depict thresholds for symmetric and asymmetric notches, respectively. Right-pointing triangles illustrate conditions in which the lower edge of the notch was 0.2 $f_P$ and the upper edge was 0.4 $f_P$. Left-pointing triangles illustrate conditions in which the lower edge of the notch was 0.3 or 0.4 $f_P$ and the upper edge was 0.1 or 0.2 $f_P$. Error bars illustrate ± 1 SD.

Inferred auditory filter shapes at 500 Hz are shown in **Fig. 2.3**. Filters inferred using the short, 30-ms maskers were broader than those inferred using the longer, 400-ms maskers. The pattern was broadly consistent across participants, although the effect of masker duration was greatest for S4. The mean ERB across participants was larger for the short than for the long masker (126 Hz vs 55 Hz, Wilcoxon test Z=28, p=0.016). In addition, the group mean ERB was consistent with the ERB of the auditory filter inferred from the mean NN curve, shown in **Fig. 2.3** (short maskers: mean ERB =126 Hz vs. ERB of mean NN curve =123 Hz; long maskers: mean ERB=55 Hz vs. ERB of mean NN curve=55 Hz). Lastly, although the fit was poorer for the short (RMSE=2.6 dB) than for the long masker (RMSE=1.4 dB), the difference was not statistically significant (Wilcoxon test Z=19, p=0.469).

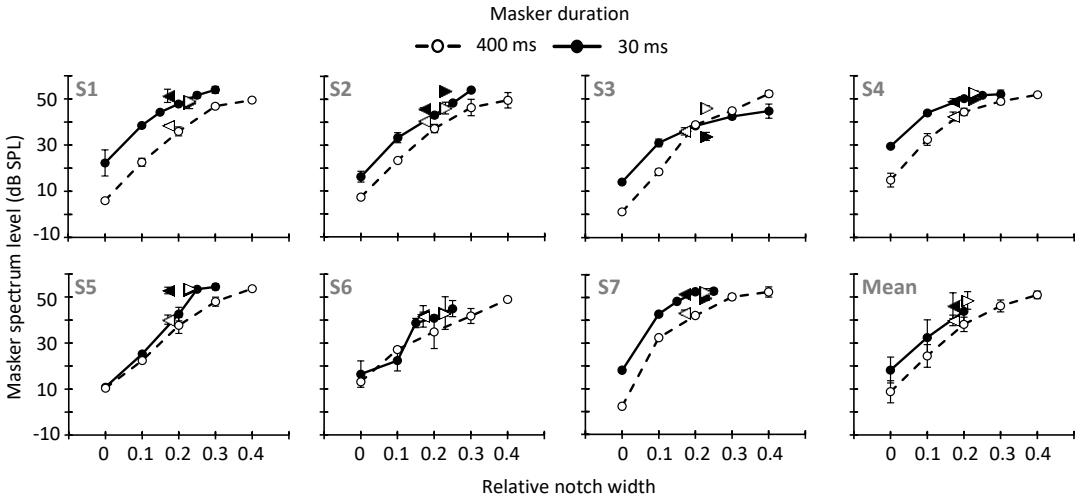


**Figure 2.2.** As in **Fig. 2.1**, but for a probe frequency of 4 kHz.

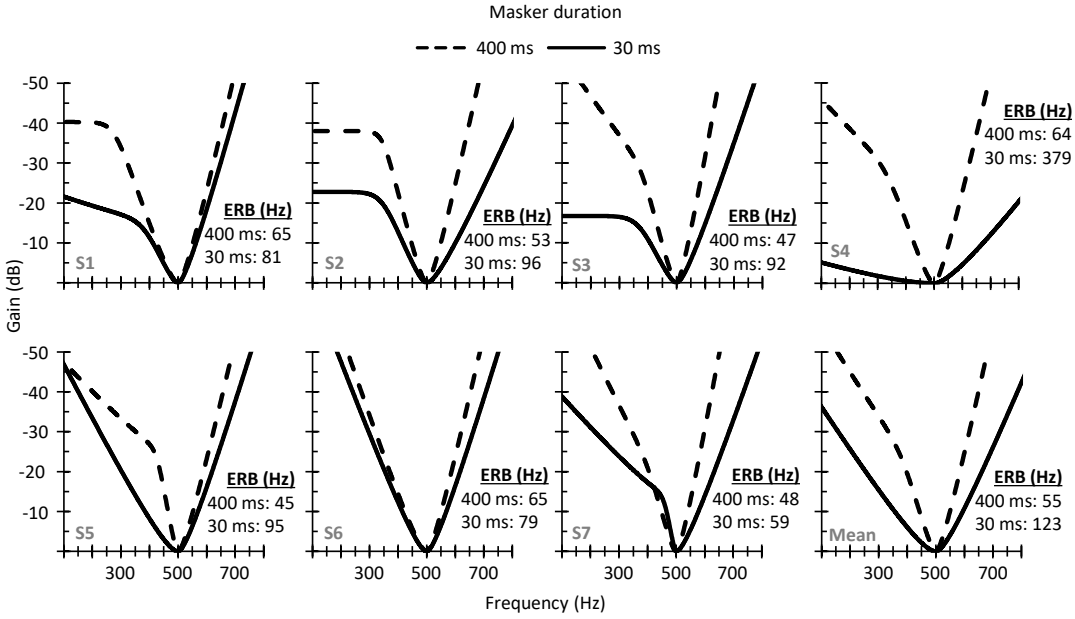


**Figure 2.3.** Auditory filter shapes [roex(*pwtp*)] centered at 500 Hz, derived from the individual data for each listener or the mean, as indicated in the bottom-left corner of each panel. Solid and dashed lines depict shapes inferred with short (30 ms) and long (400 ms) maskers, respectively. Equivalent rectangular bandwidth (ERB) estimates are shown within each panel for each masker duration.

**Table 2.1**. roex (*pwtp*) parameters, the detector efficiency (*K*) or signal-to-noise ratio at the output of the filter, and RMSE for each participant and condition for a probe frequency of 500 Hz. Note that abnormally large *t* values result in filter shapes with a horizontal low-frequency tail at the height determined by *w*. These *t* values should not be taken literally since any sufficiently large value of *t* would give the same filter shape.

| Participant | Masker duration (ms) | $p_l$ | 10*log(*w*) | *t* | $p_u$ | *K* | RMSE (dB) |
|---|---|---|---|---|---|---|---|
| S1 | 30 | 24.1 | -14.8 | 6.57 | 31.0 | 0.80 | 1.6 |
| | 400 | 26.3 | -40.2 | 1719268.00 | 37.3 | 7.87 | 2.5 |
| S2 | 30 | 23.9 | -22.8 | 2203207.00 | 19.4 | -3.61 | 2.9 |
| | 400 | 36.6 | -38.0 | 32815954.00 | 39.3 | 10.10 | 0.9 |
| S3 | 30 | 22.4 | -16.7 | 47071839.00 | 26.3 | -9.67 | 1.6 |
| | 400 | 38.3 | -22.5 | 2.95 | 48.5 | 12.20 | 2.2 |
| S4 | 30 | 3.03 | -466 | 293.00 | 11.2 | 0.56 | 1.5 |
| | 400 | 27.7 | -22.0 | 2.89 | 36.8 | 11.50 | 1.2 |
| S5 | 30 | 16.8 | -171.5 | 7032.00 | 27.9 | -5.35 | 2.2 |
| | 400 | 56.2 | -23.6 | 5.90 | 37.3 | 9.61 | 1.1 |
| S6 | 30 | 22.7 | -180 | 11.40 | 28.4 | 13.10 | 5.9 |
| | 400 | 25.7 | -194 | 367.00 | 38.3 | 11.80 | 0.5 |
| S7 | 30 | 66.3 | -13.3 | 6.56 | 24.9 | -3.85 | 2.4 |
| | 400 | 39.5 | -13.6 | 2.13 | 46.8 | 9.72 | 3.3 |
| Mean | 30 | 13.5 | -122 | 1.02 | 20.6 | -2.21 | 0.9 |
| | 400 | 34.9 | -17 | 2.44 | 39.4 | 11.10 | 1.3 |

**Table 2.2.** As **Table 2.1** but for a probe frequency of 4 kHz.

| Participant | Masker duration (ms) | $p_l$ | 10*log(*w*) | *t* | $p_u$ | *K* | RMSE (dB) |
|---|---|---|---|---|---|---|---|
| S1 | 30 | 47.3 | -23.3 | 3.01 | 64.1 | -19.60 | 0.9 |
| | 400 | 44.6 | -39.0 | 3.89 | 66.6 | -3.70 | 0.8 |
| S2 | 30 | 91.1 | -12.9 | 3.28 | 49.0 | -9.30 | 0.7 |
| | 400 | 46.4 | -32.8 | 3.29 | 50.9 | -1.65 | 0.7 |
| S3 | 30 | 71.3 | -20.3 | 5.31 | 30.8 | -17.40 | 2.6 |
| | 400 | 54.1 | -29.2 | 2.56 | 58.5 | -1.83 | 0.9 |
| S4 | 30 | 50.2 | -18.4 | 3.87 | 57.6 | -14.60 | 0.4 |
| | 400 | 52.0 | -24.6 | 3.12 | 86.7 | 0.78 | 0.7 |
| S5 | 30 | 52.8 | -57.8 | 8603551.00 | 116.0 | -8.74 | 3.0 |
| | 400 | 42.5 | -31.7 | 2.65 | 96.8 | -9.82 | 1.0 |
| S6 | 30 | 40.7 | -39.3 | 11.9 | 41.1 | -0.53 | 2.6 |
| | 400 | 61.1 | -18.7 | 3.39 | 38.6 | 1.80 | 1.6 |
| S7 | 30 | 73.0 | -30.4 | 4.67 | 80.3 | -18.10 | 0.5 |
| | 400 | 123.0 | -36.6 | 6.78 | 78.9 | -1.18 | 1.0 |
| Mean | 30 | - | - | - | - | - | - |
| | 400 | 46.3 | -26.4 | 2.64 | 53.6 | -3.88 | 0.4 |

Inferred auditory filter shapes at 4 kHz are shown in **Fig. 2.4**. For the "mean listener," a filter is plotted only for the long masker. This is because for the short masker thresholds could be measured only for the 0, 0.1, 0.2 and (0.2 and 0.4) notch widths, so there were not enough data to fit the roex to the corresponding mean NN curve. The group mean ERB was not statistically different for the short and the long masker (294 Hz vs. 277 Hz; F(1,6)=0.56, p=0.482). On average, roex(*pwtp*) fits were poorer for short (RMSE=1.7 dB) than for long maskers (RMSE=0.96 dB), but the difference was not statistically significant [paired Student's *t* test t(6)=1.57, p=0.167].

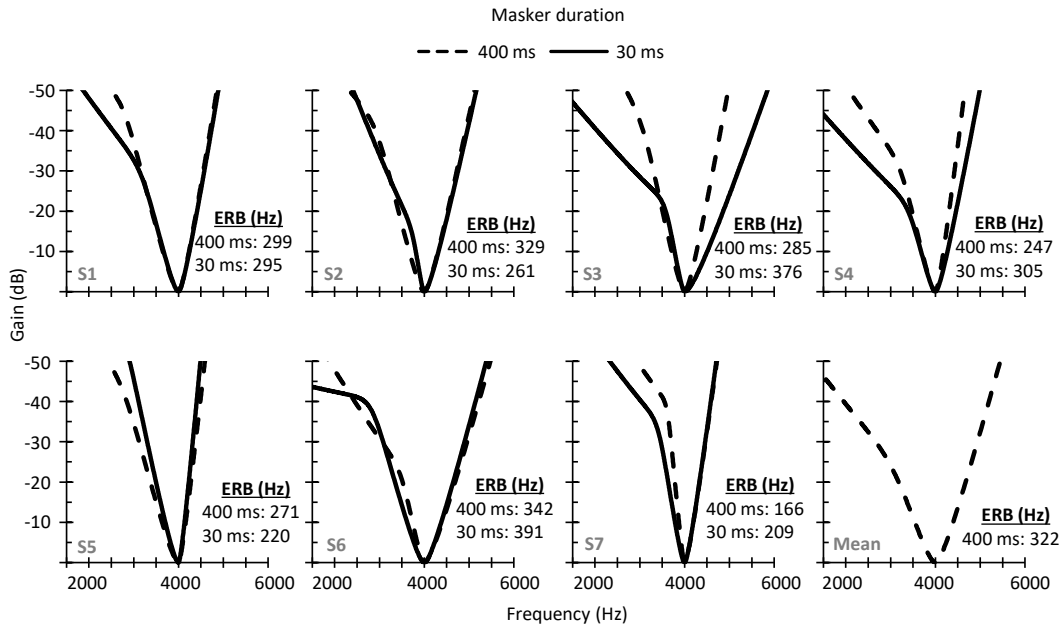


**Figure 2.4.** As **Fig. 2.3**, but for the probe frequency of 4 kHz. Note that for the "mean listener" (bottom right panel), the auditory filter is only shown for the long (400 ms) masker. This is because thresholds for the short (30 ms) masker could not be measured for all notch widths, hence there were not enough data points to fit the roex curve properly.

## 2.4. DISCUSSION

The aim of the present study was to investigate a potential effect of masker duration on estimates of human cochlear tuning obtained using the forward masking method of Shera et al. (2002). The method was applied for two masker durations: one chosen to be short enough (30 ms) not to activate the MEMR or the MOCR, and one chosen to be long enough (400 ms) that it almost certainly activated either or both reflexes when masker levels exceed the activation thresholds of the reflexes. At 4 kHz, mean ERBs were not significantly different for the two masker durations ($ERB_{30ms}$=294 Hz vs $ERB_{400ms}$=277 Hz). In contrast, at 500 Hz, the mean ERB was significantly broader for the shorter than for the longer masker ($ERB_{30ms}$=126 Hz vs $ERB_{400ms}$=55 Hz).

The effect of masker duration on behavioral estimates of frequency selectivity has been scarcely investigated, and the few existing related studies have yielded diverse and

inconclusive results. Bacon and Jesteadt (1987) found sharper psychoacoustical tuning curves (PTCs) at 1 kHz measured with long (400 ms) than with short (50 ms) pure tone forward maskers. Kidd et al. (1984) found opposite effects of increasing masker duration (from 30 to 300 ms) on the tuning of PTCs at 3 kHz for two participants with normal hearing. Jennings and Strickland (2012a) inferred tuning at 4 kHz using short (20 ms) pure-tone and NN maskers and compared their estimates with those of previous studies that used longer maskers. They concluded that masker duration may not influence estimates of frequency selectivity at 4 kHz. Our findings at 4 kHz are thus consistent with those of Jennings and Strickland (2012a). We cannot compare our findings at 500 Hz with previous studies because, as far as we know, this is the first study to investigate the effect of masker duration on auditory filter shape at 500 Hz.

### *2.4.1. Comparisons with previous human cochlear tuning estimates*

As commented in **Chapter 1**, several equations have been proposed to describe human auditory frequency selectivity (inferred using NN maskers) as a function of frequency. **Figure 2.5** compares the present $Q_{ERB}$ values for short and long maskers (squares and triangles, respectively) with predictions made using three such equations: (1) Glasberg and Moore (1990) (GM90); (2) Oxenham and Shera (2003) (OS03), which was restricted to frequencies ≥1 kHz; and (3) Leschke et al. (2022) (LRSO22), which extended the work of Oxenham and Shera (2003) to frequencies <1 kHz. Note that the present data and the data used to develop the equations on OS03 and LRSO22 were obtained using forward masking and fixed, low-level probes (i.e., the method of Shera et al., 2002), whereas GM90 is based on data obtained using simultaneous masking and fixed moderate-level maskers. At 4 kHz, the present $Q_{ERB}$ values match those of OS03 and LRSO22, and all suggest sharper tuning than GM90. At 500 Hz, our mean $Q_{ERB}$ for 400-ms maskers is roughly coincident with that of LRSO22, which was obtained using 200-ms forward maskers. However, the $Q_{ERB}$ inferred using short (30 ms) maskers is closer to that obtained in simultaneous masking (GM90) and significantly smaller than that obtained using longer, forward maskers (LRSO22).

The reason for this pattern of results is not clear. Leschke et al. (2022) discussed that forward and simultaneous NN masking procedures may yield closer $Q_{ERB}$ estimates at lower than at higher frequencies because cochlear responses are more linear in the cochlear apex than in the base (Cooper and Rhode, 1995; Versteegh et al., 2011; Recio-Spinoso and Oghalai, 2017, 2018). However, there is no clear explanation as to why the $Q_{ERB}$ values obtained using simultaneous maskers should match values inferred using short but not long forward maskers. Given that different procedures (simultaneous masking with fixed moderate-level maskers vs forward masking with fixed low-level probe) are known to result in different bandwidth estimates (Eustaquio-Martín and Lopez-Poveda, 2011; Lopez-Poveda and Eustaquio-Martín, 2013), the correspondence between the two sets of values could be coincidental.

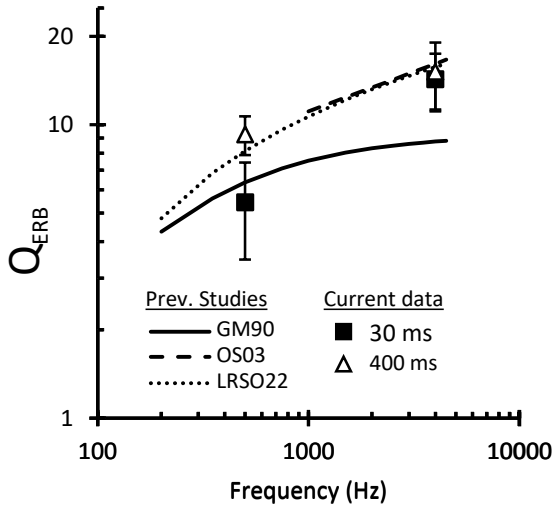


**Figure 2.5.** $Q_{ERB}$ estimates as function of frequency. $Q_{ERB} = f_P$ /ERB. Symbols and error bars illustrate the present estimates. Filled and open symbols depict values obtained using short (30 ms) and long (400 ms) maskers, respectively. Error bars illustrate ± 1 SD. Curves show regression equations reported in previous studies. GM90 (solid line): Glasberg and Moore (1990); OS03 (dashed line): Oxenham and Shera (2003); LRSO22 (dotted line): Leschke et al. (2022).

### *2.4.2. Factors that may contribute to the masker-duration effect at 500 Hz*

Several factors could be contributing to the observed effect of masker duration on auditory filter bandwidth at 500 Hz.

One potential factor is the MEMR. Rosen and Baker (1994) showed that the NN maskers used to infer auditory filters at 500 Hz and 1 kHz can activate the MEMR, and Borg et al. (1990) showed that the frequency tails of AN tuning curves correlate with MEMR activation. The MEMR can be activated by sounds longer than about 100 ms (Borg, 1982; Church and Cudahy, 1984) with overall levels higher than about 55 dB SPL (Feeney et al., 2017). Low-frequency (500 Hz and 1 kHz) ipsilateral sounds are particularly effective in activating the MEMR (McGregor et al., 2018), and MEMR activation attenuates low frequency sounds (Liberman and Guinan, 1998). Therefore, the long (400 ms) NN maskers used here could have activated the MEMR, particularly those with wider notch widths (higher levels), and their effect would be greater at 500 Hz. Because the peak of the MEMR effect occurs tens of milliseconds after the elicitor offset (Church and Cudahy, 1984), a masker-activated MEMR may have attenuated both the masker and probe approximately equally. Therefore, if the masker and the probe are processed equally linearly or compressively at the level of the basilar membrane (BM), MEMR activation would not change the masker level. If the probe is processed linearly (e.g., because its level is lower) and the masker is processed compressively, with an active MEMR, a lower masker level would be required to mask the (less effective) probe. Because the MEMR would be more likely active and stronger at higher levels (i.e., for wider notch widths), the MEMR would presumably not change or reduce masker thresholds more for the wider than for the narrower notches.

This pattern, however, would suggest equal or narrower tuning for longer than for shorter maskers, which would be opposite to the present observations (**Fig. 2.1**).

Another factor could be the MOCR. The human MOCR can be activated by ipsilateral sounds longer than about 50 ms with levels higher than about 40 dB SPL (Backus and Guinan, 2006). Noise is an effective MOCR elicitor and the MOCR effect is greater at 500 Hz than at 4 kHz (Lilaonitkul and Guinan, 2009, 2012). Therefore, the present maskers could have activated the MOCR, particularly at 500 Hz. For basal cochlear regions, MOCR activation reduces cochlear gain, resulting in overall more linear cochlear mechanical responses and broader tuning (Murugasu and Russell, 1996; Dolan et al., 1997; Cooper and Guinan, 2006). Here, tuning was narrower for the maskers that were long enough to activate the MOCR. In other words, the effect of masker duration on tuning was contrary to expectations based on a hypothetical MOCR effect. However, the characteristics of cochlear mechanical responses are different in the cochlear apex than in the base. For example, contrary to what happens in the base, cochlear tuning in the apex is sharper at higher than at lower levels (Recio-Spinoso and Oghalai, 2017, 2018). Therefore, perhaps MOCR activation has different effects on tuning in the apex than in the base.

A third possible factor is spectral splatter of the masker. The present long and short maskers were generated to have identical spectra but were different in duration. Therefore, although maskers were ramped, the excitation of the shorter (30 ms) maskers may have spread to a broader cochlear region than that of the longer (400 ms) maskers, particularly at 500 Hz. This could have made the shorter maskers less effective than the longer maskers in masking the fixed-level probe for narrow notch widths, but gradually more effective than the longer masker as the notch width increases. As a result, thresholds for the shorter maskers would be higher for narrower notch widths, but lower for the wider notches widths. The effects of spectral splatter would be greater at 500 Hz than at 4 kHz, producing a pattern broadly consistent with the present observations.

Masker-probe confusion is another factor that could have contributed to the findings at 500 Hz. In forward masking, listeners can confuse the probe with the masker when they are acoustically similar (e.g., in spectrum, duration, and/or level) (Neff, 1985, 1986). If a listener's strategy to detect the probe is to detect a different sound (the probe) at the end of the masker in each listening interval, the greater the acoustic similarity between the masker and the probe, the harder it would be to detect the probe. The short maskers were similar in duration to the probe both at 500 Hz and 4 kHz. At 500 Hz, however, the NN maskers had a bandwidth of 250 Hz, thus they were spectrally more similar to the probe than the NN maskers used at 4 kHz, whose bandwidth was 2 kHz. Furthermore, at 500 Hz, due to its short duration and reduced bandwidth short maskers contained energy fluctuations (de Boer, 1962, 1966) that could have been confused with the probe, or, conversely, the probe could have been mistaken for another modulation cycle of the masker. In summary, the similarity between probe and maskers may have made it more difficult for listeners to discriminate the target masker+probe interval from the masker only intervals. This may have made the shorter maskers appear more effective than the longer masker, as observed for larger notch widths at 500 Hz (**Fig. 2.1**).

Another factor may be that the level range was slightly different for the longer and shorter maskers, at least for some subjects (**Fig. 2.1**). Filter shape and tuning can change with level

(Ruggero et al., 1997; Robles and Ruggero, 2001; Recio-Spinoso and Oghalai, 2017, 2018) and comparing filter shapes and tuning estimates at different masker levels can result in different outcomes (Eustaquio-Martín and Lopez-Poveda, 2011; Lopez-Poveda and Eustaquio-Martín, 2013). Therefore, tuning estimates may be different for different masker durations because of differences in level, regardless of the masker duration. It might be possible to control this aspect in future studies by adjusting the time gap to equate the level of the two maskers in the no notch condition.

Although the theoretical effects of some of these factors appear more consistent with the present observations than others, it would be wrong to conclude that the factor(s) in question are responsible for observed effect of duration at 500 Hz. The factors just discussed are not mutually exclusive; they could occur simultaneously and interact in a nonlinear manner. Further research is necessary to elucidate the relative importance of these factors on the effect of masker duration.

### *2.4.3. Factors that may contribute to the masker-duration effect at 4 kHz*

At 4 kHz, masker thresholds were lower for the 400-ms than for the 30-ms maskers (**Fig. 2.2**).

Of the factors described in the preceding section, it is unlikely that the MEMR, spectral splatter or confusion affected the results at 4 kHz. This is because the MEMR is elicited by and attenuates mostly low-frequency sounds, spectral splatter is virtually negligible at higher frequencies for the present stimulus duration, and the NN maskers and probes are acoustically more dissimilar at 4 kHz. The MOCR, however, could be activated at 4 kHz by the longer but not by the shorter maskers. Activation of the MOCR would presumably inhibit the cochlear gain, which would reduce the cochlear response to the low-level probe and the maskers. If the masker was processed compressively, the reduction in probe excitation would cause a reduction in the masker level needed to mask the probe (see, e.g., Aguilar et al., 2013). Therefore, the greater effectiveness of the longer maskers at 4 kHz could be consistent with a cochlear gain reduction caused by an activation of the MOCR by the long maskers. Although previous studies have reached similar conclusions about the role of the MOCR in forward masking (Jennings et al., 2009; Jennings and Strickland, 2012b; DeRoy Milvae and Strickland, 2018, 2021), classical models of forward masking can also account the effect of masker duration without considering cochlear gain inhibition caused by MOCR activation (e.g., Dau et al., 1996).

Regardless of the mechanism behind the difference in masker threshold, the average slope of the NN curves was similar for the two masker durations, and so was the inferred tuning (**Fig 2.2** and **2.4**)

## 2.5. CONCLUSIONS

The main findings were as follows:

1) At 4 kHz, ERBs were not significantly different when inferred using short (30 ms) and long (400 ms) maskers ($ERB_{30ms}$=294 Hz vs $ERB_{400ms}$=277 Hz).

2) At 500 Hz, ERBs were significantly broader for the 30-ms than for the 400-ms masker (126 Hz vs 55 Hz).
3) Caution should be exercised when interpreting behavioral frequency selectivity in terms of cochlear mechanical tuning at low frequencies.

# 3.

# EFFECTS OF IPSILATERAL, CONTRALATERAL AND BILATERAL NOISE PRECURSORS ON HUMAN PSYCHOACOUSTICAL TUNING CURVES[2]

## 3.1. INTRODUCTION

As discussed in **Chapter 1**, the auditory system can adapt to noisy environments by changing the way it encodes and processes acoustic spectral and temporal information (Willmore and King, 2023). The medial olivocochlear reflex (MOCR) is thought to be one of the physiological mechanisms involved in adaptation to noise (Marrufo-Pérez and Lopez-Poveda, 2022). When medial olivocochlear neurons are reflexively activated by ipsilateral or contralateral sounds, they inhibit the gain of the cochlear amplifier, linearize basilar membrane (BM) responses, and modify cochlear tuning (Guinan 2006, 2010; Lopez-Poveda, 2018). Since behavioral auditory frequency selectivity is thought to be mainly determined by the frequency tuning of auditory nerve (AN) neurons (Fletcher, 1940; Evans and Wilson, 1973; Evans, 2001), and this is largely determined by the tuning of cochlear mechanical responses (Narayan et al., 1998; but see Recio-Spinoso and Oghalai, 2017), the MOCR may change auditory frequency selectivity in noise. The aim of the present study was to investigate the effect of MOCR activation on human auditory frequency selectivity at 500 Hz and 4 kHz when the MOCR is elicited by ipsilateral, contralateral, and bilateral broadband noise.

Previous human studies have employed different methods to elicit the MOCR during the estimation of frequency selectivity [reviewed by Lopez-Poveda (2018) and Jennings (2021)]. Some studies have measured psychoacoustical tuning curves (PTCs) while eliciting the MOCR using contralateral acoustic stimulation (CAS) (Kawase et al., 2000; Quaranta et al., 2005; Vinay and Moore, 2008; Aguilar et al., 2013; Wicher, 2013; Wicher and Moore, 2014). Others have

[2] The work described in this chapter is presented in the following publication: López-Ramos, D., López-Bascuas, L.E., Eustaquio-Martín, A. and Lopez-Poveda, E.A. (under review). Effects of ipsilateral, contralateral, and bilateral noise precursors on psychoacoustical tuning curves in humans. *Hearing Research*. Submitted 13 June 2024.

used ipsilateral precursor sounds to elicit the MOCR while measuring frequency selectivity using notched-noise tuning curve (NNTCs) (Strickland, 2001; Jennings and Strickland, 2012b) or PTCs (Jennings et al., 2009; Jennings and Strickland, 2012b).

However, methodological issues in previous studies preclude drawing firm conclusions regarding the effects of the MOCR on auditory frequency selectivity. In some studies, for example, the CAS used to elicit the MOCR was noise and was presented *simultaneously* with the masker and probe sounds employed to measure frequency selectivity (for narrowband noise elicitors: Quaranta et al., 2005; Vinay and Moore, 2008; Wicher and Moore, 2014; for broadband noise elicitors: Kawase et al., 2000; Aguilar et al., 2013; Wicher, 2013; Wicher and Moore, 2014). In other studies, the MOCR was elicited by ipsilateral pure tones (Jennings et al., 2009; Jennings and Strickland, 2012b) or broadband noise (Strickland, 2001) presented *before* the masker and probe. This was intended to prevent the MOCR elicitor sound from suppressing the cochlear response to the masker or the probe. These approaches may produce different results because of the different type (tones vs noise), laterality (ipsilateral vs contralateral), and presentation time (before vs simultaneous) of the MOCR eliciting sounds relative to the stimuli used to measure frequency selectivity. In addition, the previous studies disregarded the potential effect of masker duration. As discussed in **Chapter 2**, it takes about 200-300 ms for the middle-ear muscle reflex (MEMR) or the MOCR to be fully activated by sound (Church and Cudahy, 1984; Backus and Guinan, 2006). Therefore, if the maskers employed to measure frequency selectivity were longer than 200-300 ms, they could trigger the MOCR or the MEMR by themselves, making it difficult to isolate the effect of masker duration from that of the precursor or CAS.

Another open issue is whether MOCR effects on frequency selectivity differ at low (<1 kHz) and high frequencies (≥1 kHz). Several physiological (Francis and Guinan, 2010; Lilaonitkul and Guinan, 2012; Xing and Gong, 2017) and psychoacoustical (Aguilar et al., 2013) studies have shown evidence that the contralateral MOCR affects frequency selectivity more at lower than at higher frequencies, while other studies have shown the opposite (Vinay and Moore, 2008). To the best of our knowledge, no study has yet measured the effects of the ipsilateral MOCR on frequency selectivity at frequencies <1 kHz.

Lastly, psychophysical evidence suggests that contralateral MOCR sound elicitors cause a smaller cochlear gain reduction than ipsilateral ones (Salloom and Strickland, 2021). Since cochlear tuning is closely linked to cochlear gain (Ruggero et al., 1997), this suggests that ipsilateral MOCR elicitors could change frequency selectivity more than contralateral ones. However, it remains to be investigated whether the MOCR affects frequency selectivity differently in the same listener depending on the laterality of the MOCR eliciting sound.

The main aim of the present study was to investigate the effect of MOCR activation on human frequency selectivity using stimuli and methods designed to minimize the confounding effects described earlier. Another aim was to compare the effect of the MOCR on frequency selectivity at low at high frequencies. A third aim was to compare the strength of the MOCR evoked by ipsilateral, contralateral, and bilateral sounds.

To address these aims, PTCs were measured in forward masking for the same listeners for probe signals of 500 Hz and 4 kHz in the presence and in the absence of ipsilateral,

contralateral, and bilateral broadband precursor noises. Although broadband noises are effective in activating the MOCR (Lilaonitkul and Guinan, 2009a) and MEMR (Church and Cudahy, 1984), here, the level of the precursor noises was set at 60 dB SPL for them to activate the MOCR with minimal activation of the MEMR (Lilaonitkul and Guinan, 2009a, 2009b; Aguilar et al., 2013). The masker and the probe were short enough (**Fig. 3.1**) to minimize the possibility that they activated the MOCR by themselves. The precursors, however, were long enough to activate the ipsilateral, contralateral, and bilateral MOCR (Backus and Guinan, 2006). Therefore, it is assumed that, by design, any effect of the precursor on the PTCs was mostly due to the activation of the MOCR and minimally due to the activation of the MEMR. To aid the interpretation of the results under this assumption, a physiologically inspired phenomenological computer model of forward masking with efferent cochlear gain control (Aguilar et al., 2013) was used to simulate the effects of the noise precursors on the PTCs.

## 3.2. MATERIAL AND METHODS

### *3.2.1. Participants*

Three men (S1, S2, S3) and one woman (S4) participated in the experiments. Their ages were 24 (S1), 24 (S2), 26 (S3), and 21 (S4) years. All of them self-reported having normal hearing and no history of hearing disorders. All participants were tested on their right ears.

### *3.2.2. Stimuli*

PTCs at 500 Hz and 4 kHz were measured using forward masking (**Fig. 3.1**). The probes were sinusoids with a duration of 10 ms (including 5-ms cosine-squared rise/fall ramps, no steady-state portion) and a level of 10 dB above the individual absolute threshold for the probe (or 10 dB SL). The use of fixed, low-level, sinusoidal probes was intended to make the cochlear region excited by the probe as narrow as possible and fixed across conditions. This maximized the chance that listeners detected the probe using the cochlear filter centered on the probe frequency, that is, it reduced off-frequency listening (Nelson et al., 2001).

The maskers were sinusoids with a duration of 30 ms (including 5-ms cosine-squared rise/fall ramps and a 20-ms steady-state portion). For the 4 kHz probe, masker frequencies were 2.0, 2.8, 3.6, 3.8, 4.0, 4.2, 4.4, and 4.8 kHz; for the 500 Hz probe, masker frequencies were 250, 350, 450, 475, 500, 525, 550, 600, 700, and 800 Hz. The time gap between the masker offset and the probe onset was 2 ms.

PTCs were measured with and without a 300-ms noise precursor (including 50-ms cosine-squared rise/fall ramps). The noise precursor had a bandwidth from 100 Hz to 18 kHz and an overall level of 60 dB SPL. No time gap was left between the precursor offset and the masker onset. Note that the total stimulus duration was 342 ms (300-ms precursor + 30-ms masker + 2-ms gap + 10-ms probe).

PTCs were measured without and with contralateral, ipsilateral, and bilateral precursors, i.e., four conditions in total. All participants completed the four conditions (repeated measures design).

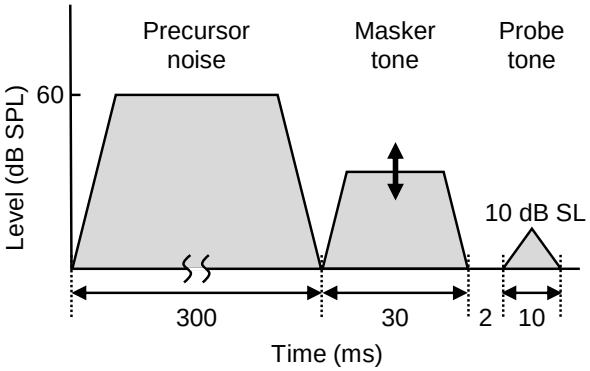


**Figure 3.1.** Representation of the stimuli used in the experiment. The masker and probe tones were delivered to the right ear and the precursor was presented contralaterally, ipsilateral, and bilaterally.

### *3.2.3. Procedure*

A two-interval, two-alternative, forced-choice adaptive procedure with feedback was employed to measure the masker level at the probe detection threshold. The initial masker level was set below the estimated threshold, and its level was changed according to a two-up, one-down adaptive procedure to estimate the 71% point on the psychometric function (Levitt, 1971). The initial masker level step was 4 dB for the first two reversals, and 2 dB thereafter. The adaptive procedure was continued until ten reversals in level were recorded. The mean masker level at the last eight reversals was regarded as the threshold. A threshold estimate was discarded when the standard deviation (SD) exceeded 6 dB. Each masker threshold was estimated three times. If the SD of the three estimates exceeded 6 dB, an additional estimate was obtained and included in the mean. Therefore, the reported masker thresholds are the mean of at least three valid threshold estimates.

Absolute hearing thresholds for the probes were estimated using a two-down, one-up adaptive rule. The absolute thresholds for the 500 Hz probe were 26 (S1), 27 (S2), 26 (S3), and 27 (S4) dB SPL. The absolute thresholds for the 4 kHz probe were 14 (S1), 7 (S2), 13 (S3), and 17 (S4) dB SPL.

### *3.2.4. Test sessions*

The complete experiment took seven two-hour sessions. Breaks were given in each session, as required. The first session was devoted to training the participants in the forward masking task and to measuring their probe absolute thresholds. Three sessions were devoted to measuring PTCs with 500-Hz probes and three sessions to measure PTCs with 4-kHz probes. The 4-kHz and 500-Hz sessions were interleaved, and the order was balanced across participants. Masker frequencies were presented in random order across participants and sessions.

### 3.2.5. Equipment

Stimuli were generated using custom-made MATLAB software (R2014a, The Mathworks, Inc.) and played via an RME Fireface 400 soundcard at a sampling rate of 44.1 kHz, and with 24-bit resolution. All stimuli were presented to the listeners using Sennheiser HD265 headphones. Listeners were seated in a sound-attenuating booth during all measurements. Sound pressure levels were calibrated by placing the headphones on a KEMAR (Knowles Electronics) equipped with a Zwislocki DB-100 artificial ear connected to a sound level meter (Brüel and Kjaer 2238). Calibration was performed at 1 kHz and the obtained sensitivity was used at all other frequencies.

### 3.2.6. Estimating PTC bandwidths

Each side of a PTC was independently fitted with a roex($pwt$) as follows (Patterson and Nimmo-Smith, 1980; Patterson et al., 1982) (**Eq. (3.1)**):

$$W(g) = -10\log_{10}\left[(1-w)(1+pg)e^{-pg} + w\left(1+\frac{pg}{t}\right)e^{-\frac{pg}{t}}\right] \quad (3.1)$$

where $g$ is the masker frequency relative to the probe frequency. $p$, $t$, determine the slopes of the PTC at its tip and tail, respectively; and $w$ is the intersection of these two slopes. $p$, $t$ and $w$ were regarded as free parameters and adjusted during the fitting. In all cases, there were fewer free parameters than data points. The purpose of the fits was to provide a smoother representation of the PTC that allowed better inference of the various PTC bandwidth measures. 3-dB-down ($BW_{3dB}$), 10-dB-down ($BW_{10dB}$), and equivalent rectangular bandwidth (ERB) estimates were obtained from the fits. The pattern of results was quite similar for the three bandwidth measures. Therefore, only ERB values and analyses are reported below.

### 3.2.7. Statistical analysis

Statistical analyses were performed using the *rstatix* (Kassambara, 2021) package in R (version 4.1.0). Shapiro-Wilk test of normality revealed that the ERB for 500-Hz and 4-kHz PTCs was normally distributed. To assess the statistical significance of the precursor effect on the ERB, a one-way repeated measures analysis of variance (RMANOVA) was performed for each probe frequency separately, having the precursor condition (none, ipsilateral, contralateral, and bilateral) as the factor. Tukey's honestly significant different (HSD) test was applied for post-hoc pairwise comparisons using the *multcomp* package (Hothorn et al., 2008).

### 3.2.8. Computer model

A physiologically inspired computer model of forward masking with efferent cochlear gain control was used to investigate if the effects of the precursor noises on the PTCs were consistent with a precursor activation of the MOCR. The model was the same as that used by Aguilar et al. (2013) to simulate the effect of CAS on PTCs. This model was based on the BM-temporal window model of Plack et al. (2002). The first stage of the model consisted of a filter that simulates the outer ear, followed by a filter that simulates the middle ear, and a dual-

resonance nonlinear filter of the human BM response (Lopez-Poveda and Meddis, 2001) with efferent control (EDRNL) (Ferry and Meddis, 2007). This first stage simulates the mechanical response of the BM. The effects of activating the medial olivocochlear efferent on the BM response are simulated by means of a parameter (ATT) that attenuates (inhibits) the cochlear gain. The output of this first stage is passed through a squaring device, followed by a temporal window, which integrates the stimulus energy over time. The output of the temporal window is then passed through a decision device. In the decision stage, the model compares two outputs: one with the masker and the probe, and one with the masker alone. The probe is assumed to be detected when the ratio of the two outputs at the time of the probe peak exceeds a criterion value, *K* (Plack et al., 2002).

Note that, in this model, forward masking is assumed to occur from the persistence of neural activity produced by the masker (or by the masker plus the ipsilateral precursor when a precursor was present) (Oxenham, 2001). Also note that in this model MOCR effects are *not* dynamic. Instead, it was assumed that the precursor inhibited the cochlear gain by a fixed, constant amount (ATT in dB) throughout the entire duration of the stimulus, including the precursor. The implications of the latter assumption are discussed later.

The model was implemented digitally in the time domain with a sampling frequency of 44.1 kHz. The model was evaluated with the same stimuli and under the same conditions as those used in the experiments. Instead of optimizing the model parameters to quantitatively reproduce the present PTCs, we used parameters published by Aguilar et al. (2013) to simulate experimental PTCs for their subjects S1 and S2. The parameters for S1 were used to simulate the present PTCs at 500 Hz, while those of S2 were used to reproduce the present results at 4 kHz. Here, the value of *K* was set equal to 7 (500 Hz) and 8 (4 kHz) to adjust the level at the tip of the simulated PTCs to the level the tip of mean experimental PTCs. The absolute probe threshold of the model was set to be the same as that reported by Aguilar et al. (2013). The efferent attenuation parameter was adjusted for each precursor laterality to mimic the pattern of experimental data.

### 3.3. RESULTS

**Figure 3.2** and **Fig. 3.3** show individual and mean PTCs measured without and with each type of precursor for probe frequencies of 500 Hz and 4 kHz, respectively. Lines represent roex fits to the data. The pattern of results was similar across participants. Therefore, mean data were used to analyze the precursor effect as a function of masker frequency in **Fig. 3.4**.

At 500 Hz (**Fig. 3.2** and **Fig. 3.4A**), masker thresholds were lower around the tips of the PTCs with the ipsilateral and bilateral precursors than without a precursor. Contralateral precursors had a small and roughly constant effect across frequencies. A one-way ANOVA revealed a significant effect of precursor type on the ERB at 500-Hz ($F_{(3,9)}$=5.69, p=0.018). Post-hoc Tukey HSD test showed that the ERB in the control condition (139 Hz) differed significantly from the ERB estimated with ipsilateral (85 Hz, p=0.002) and bilateral (75 Hz, p<0.001) precursors. By contrast, the ERB in the control condition did not differ from the ERB estimated with a contralateral precursor (118 Hz, p=0.49).

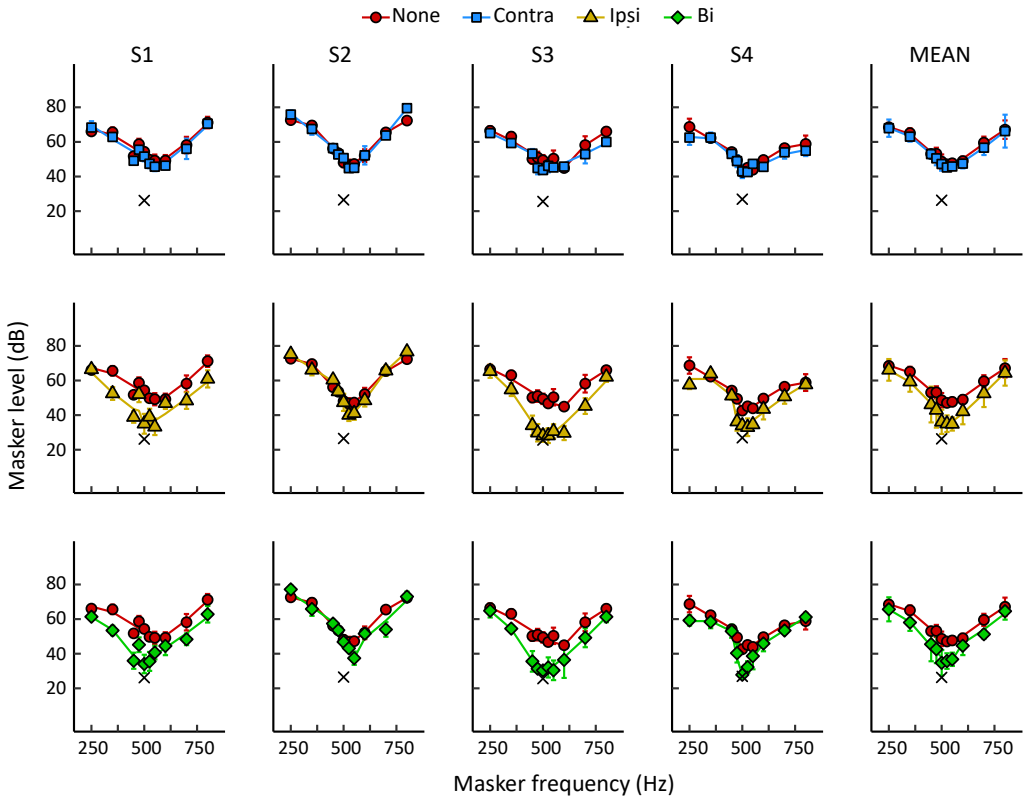


**Figure 3.2.** Effect of precursor noise laterality on psychoacoustical tuning curves (PTCs) for a probe frequency of 500 Hz. Each column is for a different subject or the mean, as indicated at the top of the column. Symbols illustrate experimental data (mean of at least three measures), and lines illustrate roex fits. Different symbols illustrate masker thresholds for different precursor literalities as indicated by the legend. Individual and mean probe levels are represented by crosses. Error bars illustrate ± 1 SD.

At 4 kHz (**Fig. 3.3** and **Fig. 3.4B**), ipsilateral and bilateral precursors had little effect on thresholds for masker frequencies near the probe frequency, but they caused a decrease in thresholds for masker frequencies far from the probe frequency, resulting in broader PTCs. The effect of contralateral precursors was much smaller, almost negligible. A one-way ANOVA revealed a significant effect of precursor type on the ERB at 4 kHz ($F_{(3,9)}$=7.24, p=0.009). Post-hoc Tukey HSD showed that the ERB in the control condition (210 Hz) differed significantly from the ERB estimated with ipsilateral (394 Hz, p<0.001) and bilateral (392 Hz, p<0.001) precursors. By contrast, ERB in the control condition did not differ from the ERB estimated with contralateral precursor (267 Hz, p=0.52).

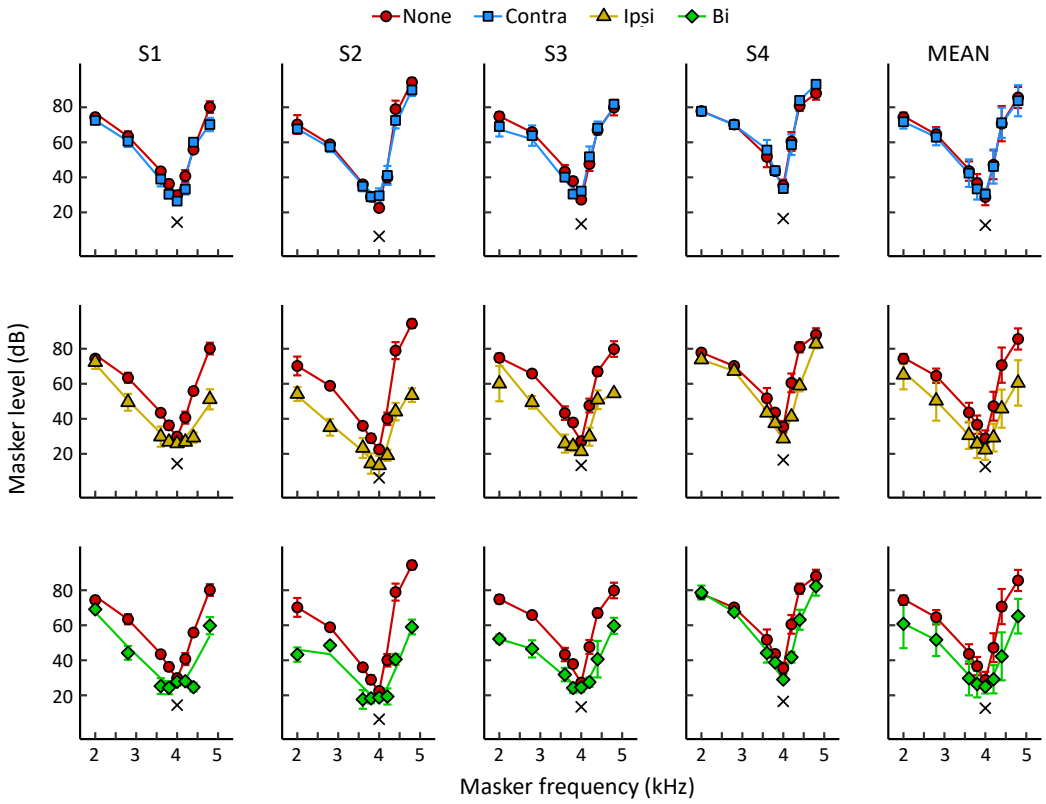


**Figure 3.3.** As **Fig. 3.2**, but for a probe frequency of 4 kHz

In summary, the pattern of the precursor effect on the PTCs was different at the two probe frequencies. At 500 Hz, ipsilateral and bilateral precursors *narrowed* the PTCs by causing a downward shift of the PTC tip with little change in thresholds for masker frequencies far from the probe frequency (**Fig. 3.4A**). At 4 kHz, the ipsilateral and bilateral precursors *broadened* the PTCs by reducing the thresholds for masker frequencies far from the probe frequency, with minimal or no change at the PTC tip (**Fig. 3.4B**).

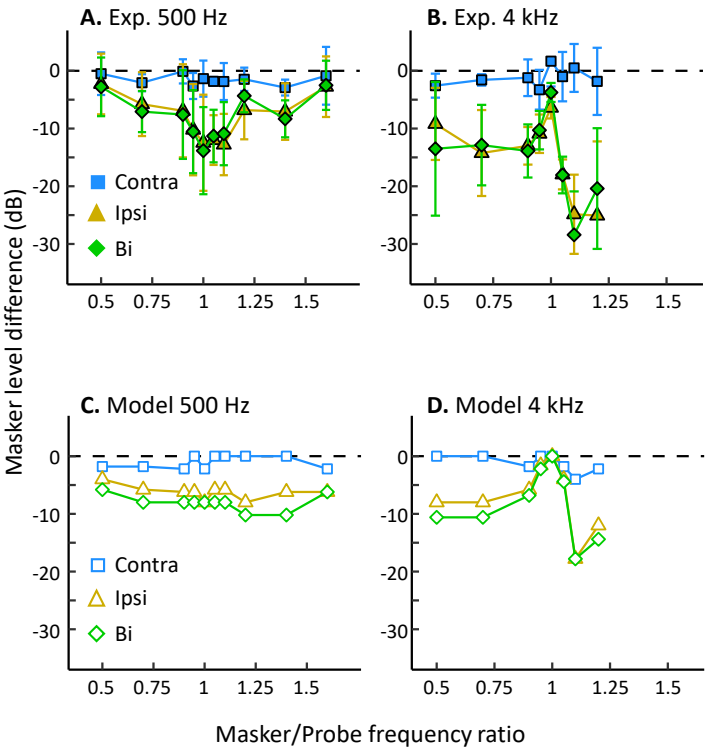


**Figure 3.4.** The effect of precursor laterality on masker threshold as a function of masker frequency in the experimental (top) and model (bottom) psychoacoustical tuning curves (PTCs) at 500 Hz (left) and at 4 kHz (right). The y axes show the difference between the group mean masker levels in each precursor condition minus the mean masker level measured without a precursor. Error bars illustrate ± 1 SD. Different symbols illustrate different precursor lateralities as indicated by the legend.

### *3.3.1. Computer model simulations*

The masker and probe used to measure the PTCs were designed for them to minimally activate the MOCR (or MEMR), while the noise precursors were designed for them to activate the MOCR (with minimal activation of the MEMR). Therefore, we assumed that, by design, the noise precursors would change cochlear tuning, thus the PTCs, by their activating the MOCR. To validate this assumption and interpret the results, we used a computational phenomenological model of forward masking with efferent control of cochlear gain (Aguilar et al., 2013). If the assumption was correct, the model should be able to mimic the precursor effects by a decrease in cochlear gain (akin to efferent inhibition). Therefore, control PTCs (without precursors) were simulated by setting the efferent ATT parameter of the model equal to 0 dB, while the effect of the various precursors on the PTCs was simulated by setting the efferent ATT parameter to a negative value (in dB) as appropriate (see **Section 3.2.8**).

**Figure 3.5** shows the simulated PTCs at 500 Hz (top row) and at 4 kHz (bottom row). Recall that the filter parameters were taken "off the shelf" (from Aguilar et al., 2013), i.e., they were not optimized to fit the experimental data. Experimental PTCs could be affected by off-frequency listening (O'Loughlin and Moore 1981), especially with the brief 500 Hz probes used here which can cause energy splatter (see, e.g., Eustaquio-Martín and Lopez-Poveda, 2011). This possibility

was accounted for in the original version of the model. However, the simulated PTCs considering off-frequency were too narrow compared to the present experimental PTCs. Therefore, the simulated 500-Hz PTCs shown in **Fig. 3.5** were simulated disregarding off-frequency listening. This will be further discussed later.

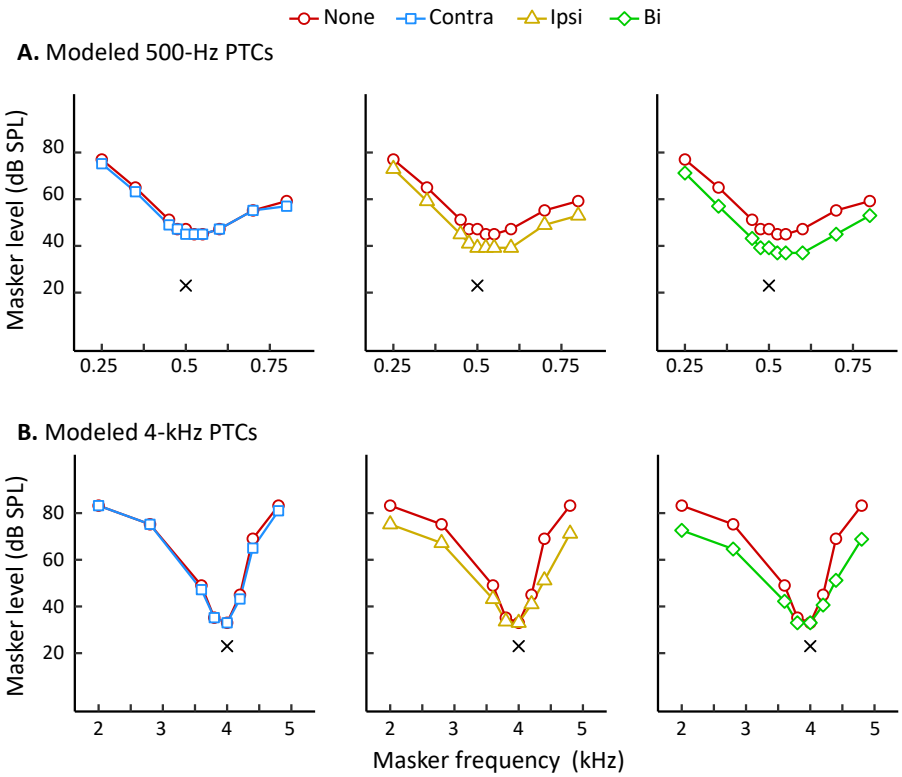


**Figure 3.5.** Model psychoacoustical tuning cuves (PTCs) at 500 Hz (**A**) and 4 kHz (**B**) with values of the efferent ATT parameter adjusted to roughly mimic the experimental effect of ipsilateral, contralateral and bilateral precursors on the PTCs. Mean probe levels are represented by crosses. See the main text for details.

At 500 Hz (**Fig. 3.5A** and **Fig. 3.4C**), setting the efferent ATT parameter equal to −1 dB caused a slight and approximately constant reduction of masker thresholds across frequencies, which was qualitatively and quantitatively consistent with the empirical effect of contralateral precursors on the PTCs. Setting ATT equal to −5 dB and −6 dB was sufficient to roughly model the experimental effect of ipsilateral and bilateral precursors, respectively, on the PTCs (we assumed that the attenuation effect of ipsilateral and contralateral precursors adds up). That is, the model produced a comparatively greater downward shift in the tip than in the tails of the PTC, as observed experimentally with ipsilateral and bilateral precursors. However, the shift was smaller in the model than in the experimental data.

At 4 kHz (**Fig. 3.5B** and **Fig. 3.4D**), setting ATT equal to −2 dB served to simulate the effect of contralateral precursors on the PTC, i.e., this attenuation was sufficient to cause a very slight broadening in the simulated PTC that was qualitatively and quantitatively consistent with the experimental PTCs measured with contralateral precursors. By setting the efferent ATT equal to −9 dB and −11 dB, respectively, the model reproduced the pattern of masker thresholds in

the 4-kHz PTCs with ipsilateral and bilateral precursors; namely, a broadening of the PTCs because of a greater reduction in masker thresholds in the tails than in the tip of the PTCs.

In summary, the model could reproduce the different experimental pattern of precursor effects at 500 Hz and 4 kHz (**Fig. 3.4**). This suggests that the observed precursor effects may be due to an inhibition of cochlear gain caused by a precursor evoked MOCR.

To better understand how MOCR inhibition could produce seemingly different effects on the PTCs at 500 Hz and 4 kHz, **Fig. 3.6** illustrates the effects of efferent attenuation on the model BM input/ouput (I/O) curves at 500 Hz (top) and at 4 kHz (bottom). The shown efferent attenuation corresponds to bilateral precursors, i.e., ATT was set equal to −6 dB and −11 dB for at 500 Hz and 4 kHz, respectively. It was assumed that the probe/masker BM response ratio at the masking threshold is equal without (p/m) as with efferent attenuation (p'/m') (Aguilar et al., 2013).

The simulations at 500 Hz will be analyzed first (**Fig. 3.6**, top row). Without efferent attenuation, the BM response produced by the on-frequency masker (m in panel A) falls in the compressive region of the I/O curve. With efferent attenuation, the BM response produced by the probe is reduced and so, with a fixed probe/masker cochlear response ratio, the masker response at threshold is lower (m' < m). Because the masker response is in the compressive region of the I/O function, the reduction in the masker response needed to meet the masking criterion makes the masker level much lower, producing a downward shift of the PTC tip. In contrast, low and high off-frequency maskers (panels B and C) are processed linearly. Therefore, the decrease in level for low and high off-frequency maskers to mask the efferent attenuated probe response is less than for the on-frequency masker.

At 4 kHz (**Fig. 3.6**, bottom row), the probe and the on-frequency masker (panel D) are both processed linearly and are equally affected by the efferent attenuation. Hence, the masker level required to mask the probe is roughly equal with and without efferent attenuation ($L_{M'} \sim L_M$). For a low off-frequency masker (panel E), the BM response is linear and not affected by efferent attenuation. Therefore, the level of the masker is lower with than without the efferent attenuation ($L_{M'} < L_M$). For a high off-frequency masker (panel F), the BM response is slightly compressive, hence the greater decrease in masker level with efferent attenuation ($L_{M'} < L_M$).

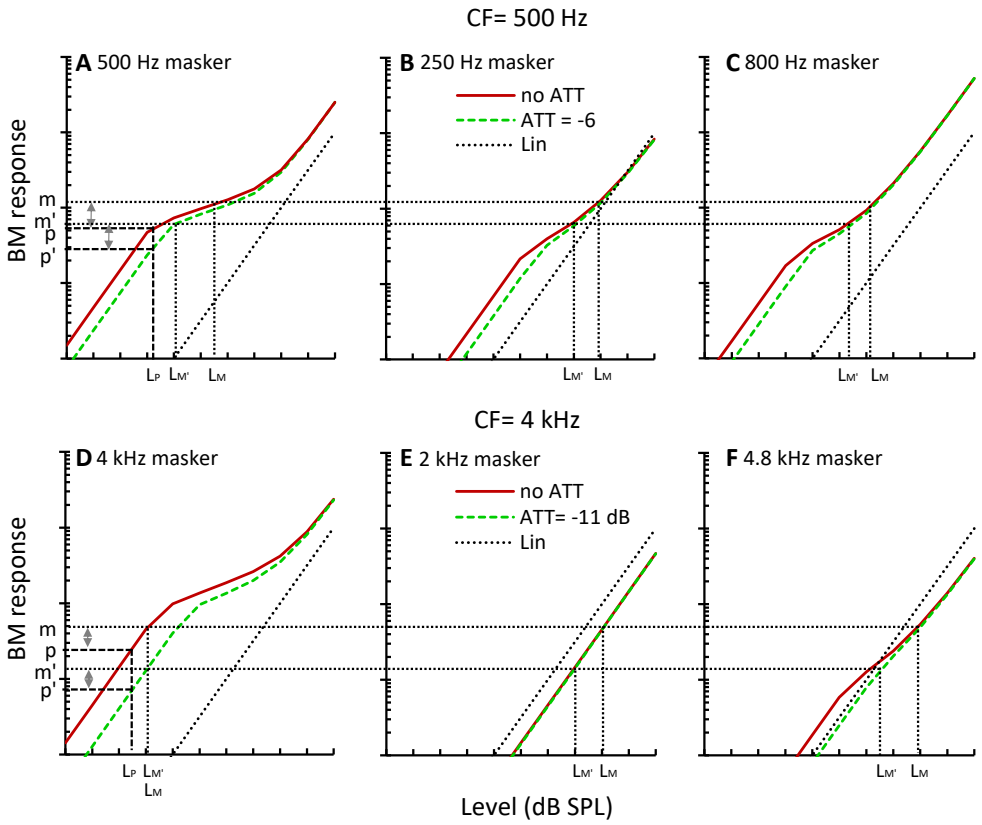


**Figure 3.6.** Model basilar membrane (BM) input/output (I/O) curves at 500 Hz (top) and 4 kHz (bottom) with and without efferent attenuation. Each panel is for a different stimulus (masker) frequency, as indicated at the top. In each panel, the green continuous line represents the BM response without efferent attenuation, the red dashed line represents the BM response with efferent attenuation, and the dotted line is a linear reference (slope = 1 dB/dB). $L_P$ denotes probe sound level; $L_M$ and $L_{M'}$ denote the masker level at the masking threshold without and with efferent attenuation, respectively. The BM response produced by the probe and the masker without and with efferent attenuation was indicated by p, p', m, and m', respectively.

## 3.4. DISCUSSION

This study investigated the effect of contralateral, ipsilateral, and bilateral broadband noise precursors on PTCs at 500 Hz and 4 kHz. At 500 Hz, ipsilateral and bilateral precursors shifted the PTC tip downward and narrowed the PTCs. In contrast, at 4 kHz, ipsilateral and bilateral precursors shifted the PTC tails downward and broadened the PTCs. Contralateral precursors had negligible effects at the two probe frequencies. The used precursors were designed (in level and duration) to trigger the MOCR with minimal activation of the MEMR (Lilaonitkul and Guinan, 2009a, 2009b; Aguilar et al., 2013) and it is unlikely that the used maskers or probes activated the MOCR due to their short duration. Therefore, it was assumed that the precursor effects are due to cochlear gain inhibition caused by MOCR activation. To aid the interpretation of these results under this assumption, a computational model of the BM response with efferent gain control was used to simulate PTCs without and with precursors. The simulated PTCs were broadly consistent with the experimental PTCs.

### 3.4.1. Comparisons with previous studies

Consistent with our results, some studies have found that the changes in cochlear frequency selectivity caused by MOCR activation may also be different across frequencies. One study that measured stimulus-frequency otoacoustic emissions (SFOAEs) suppression tuning curves found sharper tuning at 1 kHz, but broader tuning at 2 and 4 kHz when the MOCR was activated by CAS (Xing and Gong, 2017). The broadening found here at 4 kHz is consistent with previous measurements of frequency selectivity in the presence of ipsilateral precursors (Strickland, 2001; Jennings et al., 2009; Jennings and Strickland, 2012b) and CAS (Quaranta et al., 2005; Vinay and Moore, 2008). To the best of our knowledge, no study has yet measured the effects of ipsilateral precursors on low-frequency PTCs. In **Chapter 2**, it was shown that the bandwidth of the 500-Hz auditory filter is smaller when estimated using longer (400 ms) than shorter (30 ms) notched-noise (NN) maskers. The possible contributions of MOCR, MEMR, spectral splatter and masker-probe confusion effects to this result were discussed. In the present study, we also found the 500-Hz PTCs to be narrower in the presence of an ipsilateral noise precursor designed to activate the MOCR. This supports the hypothesis that the MOCR may be involved in the masker duration effect on behavioral estimates of cochlear tuning reported in **Chapter 2**.

We found contralateral precursors to have a virtually negligible effect on PTCs. This is not consistent with previous studies that reported a broadening (Aguilar et al., 2013) and a sharpening (Vinay and Moore, 2008) of 500-Hz PTC with CAS. However, unlike here, where maskers were short to prevent them from activating the MOCR, previous studies used long (Aguilar et al., 2013) or continuous (Vinay and Moore, 2008) maskers that could have activated the ipsilateral MOCR by themselves. Therefore, it is possible that previous findings may not be solely due to the contralateral MOCR.

It is difficult to infer the relative strength of the MOCR across frequencies directly from the experimental data because the precursors affected PTCs differently at low and high frequencies. However, we could approximately infer the MOCR strength for each precursor using the model. The simulations broadly matched the experimental data assuming less efferent attenuation at 500 Hz than at 4 kHz (−6 dB vs −11 dB, respectively, for bilateral precursors). This seems consistent with the higher density of MOCR terminals around 4 kHz in human young subjects (Liberman and Liberman, 2019) but inconsistent with click-evoked otoacoustic emissions (CEOAEs) and stimulus-frequency otoacoustic emissions (SFOAEs) studies that show a decreasing MOCR strength with increasing probe frequency (Francis and Guinan, 2010; Lilaonitkul and Guinan, 2012; Xing and Gong, 2017).

The change in PTC bandwidth was greater for ipsilateral than for contralateral precursors. This is consistent with previous psychophysical studies that reported less cochlear gain reduction with contralateral than with ipsilateral precursors (Salloom and Strickland, 2021). A greater gain reduction with ipsilateral than with contralateral precursors could be explained by a greater number of crossed ipsilateral efferent fibers than uncrossed contralateral fibers. Although the number of each type of fiber in humans remains unknown, data from other animals are consistent with this idea (Robertson, 1984; Liberman and Brown, 1986; Brown, 1989).

### 3.4.2. Model simulations

The model simulations suggest that the sharpening of the 500-Hz PTCs and the broadening of the 4-kHz PTCs in the presence of ipsilateral and bilateral precursors can both arise from an interaction of efferent gain inhibition (caused by the activation of the MOCR by the precursors) with compression affecting the maskers (**Fig. 3.6**).

However, the model results should be interpreted with caution. First, ipsilateral (and bilateral) precursors can affect probe detection through two processes: increased forward masking and cochlear gain inhibition by activation of the MOCR. Models that account for these two processes reproduce the experimental results better than those that only consider the masking process (Jennings et al., 2009; Jennings and Strickland, 2012b). These two processes (increased masking and efferent gain inhibition) were accounted for in the present model. That is, forward masking was greater when the ipsilateral (or bilateral) precursor was present than when it was absent because the temporal window integrated the excitation of the precursor, and efferent gain inhibition was modeled by setting the ATT parameter to a negative value (in dB). Although the experimental procedure employed here does not allow isolating the effects of forward masking and gain inhibition, the patterns of experimental and simulated PTCs suggest that the precursors affected the PTCs mainly by efferent gain inhibition. This is because if the precursors affected probe detection by forward masking, they would have reduced masker levels equally at the tip and the tails of the PTCs and this was not the case. Instead, the ipsilateral precursors affected masker thresholds differently for on- and off- frequency masker, a pattern more consistent with MOCR activation (Jennings et al., 2009; Jennings and Strickland, 2012b).

Second, our simulations were performed assuming that the efferent attenuation caused by ipsilateral and contralateral precursors add up. Previous studies have questioned the idea that the strength of the bilateral MOCR is the sum of the strength of the ipsilateral and contralateral MOCR (Berlin et al., 1995; Philibert et al., 1998; Backus and Guinan, 2006; Lilaonitkul and Guinan, 2009a, 2009b). Although the model results are roughly consistent with the experimental data, the small effects of the contralateral precursor make it difficult to clarify whether the ipsilateral and contralateral effects add up in humans.

Third, Aguilar et al. (2013) showed that CAS at 60 dB SPL did not activate the MEMR. Therefore, here it has been assumed that the used noise precursors (at 60 dB SPL) did not activate the MEMR. Although this could be true for contralateral precursors, the MEMR is more easily activated by ipsilateral than contralateral sounds (McGregor et al., 2018). Therefore, we cannot rule out that the precursors alone (or combined with the maskers) activated the MEMR.

Fourth, the model used in this study can be improved in several ways. First, by incorporating the possibility of adjusting the middle ear impedance as a result of MEMR activation. And, second, by making the activation of the MOCR and MEMR dynamic and dependent on the model's response to the acoustic stimuli in a physiologically inspired way. In this sense, a more recent version of the model (Grange et al., 2022), which incorporates these two improvements, may help in the interpretation of the present results. Other subcortical models (e.g., Farhadi et al., 2023) in which cochlear gain control from the medial olivocochlear neurons is regulated by inputs from the cochlear nucleus and inferior colliculus may be helpful in interpreting the effects of noise on frequency selectivity.

### *3.4.3. Final remark on PTCs bandwidth*

The original model accounted for off-frequency listening at 500 Hz. When the model was evaluated including off-frequency listening, the simulated PTC (not shown) was narrower than the experimental PTCs. This suggests that the experimental PTCs were barely affected by off-frequency listening. This idea is supported by the fact that at 500 Hz, the ERB inferred from the PTCs without precursor was 139 Hz close to the ERB inferred using NN maskers of similar short duration (126 Hz) (see **Chapter 2**). However, the ERB inferred using short tonal or NN maskers were broader those inferred using long-duration NN maskers (55 Hz). As discussed in **Chapter 2**, the broader ERB inferred using short maskers could be due to short low-frequency NN or tonal maskers causing energy splatter.

## 3.5. CONCLUSIONS

This study investigated the effects of ipsilateral, contralateral and biliteral noise precursors on PTCs measured with short maskers. The main findings were:

1) Contralateral precursors barely affected PTCs at 500 Hz or 4 kHz.
2) At 500 Hz, ipsilateral and bilateral precursors sharpened the PTCs by decreasing the threshold for maskers at or around the PTC tip.
3) At 4 kHz, ipsilateral and bilateral precursors broadened the PTCs by decreasing the threshold for masker frequencies in the PTC tails.
4) The pattern of results is broadly consistent with cochlear gain inhibition caused by an activation of the MOCR by the precursors.

# 4.

# ADAPTATION TO NOISE IN SPECTROTEMPORAL MODULATION DETECTION AND WORD RECOGNITION[3]

## 4.1. INTRODUCTION

Background noise can hamper hearing. The negative effect of noise, however, diminishes as the target sound is delayed a few hundred milliseconds from the noise onset. The improvement in auditory function as the target is delayed in the noise has been referred to as adaptation to noise (Cervera and Ainsworth, 2005; Marrufo-Pérez et al., 2018a, 2020) and occurs in various auditory tasks including pure tone detection (also known as "overshoot") (Elliot, 1965; Zwicker, 1965a,b; Bacon and Takahashi, 1992; Jennings et al., 2011), amplitude modulation (AM) detection (Almishaal et al., 2017; Marrufo-Pérez et al., 2018b; Wojtczak et al., 2019) and speech recognition (Cervera and Gonzalez-Alvarez, 2007; Ben-David et al., 2012; Marrufo-Pérez et al., 2018a, 2020) (reviewed by Marrufo-Pérez and Lopez-Poveda, 2022). Regarding speech recognition, normal-hearing (NH) listeners can recognize up to 30% more syllables and tolerate 3 dB more noise while maintaining 50% correct word recognition when the speech signal is delayed ~300 ms from the noise onset (Cervera and Gonzalez-Alvarez, 2007; Ben-David et al., 2012; Marrufo-Pérez et al., 2018a). It remains unknown which of the myriad speech cues improve as the speech is delayed in the noise. In this chapter, it is investigated if adaptation in speech recognition relates to improvements in the encoding of spectral (SM), temporal (TM), and/or spectrotemporal modulations (STM).

As commented in **Chapter 1,** adaptation to noise could be caused by a noise-induced shift of the dynamic range of auditory neurons (reviewed in Marrufo-Pérez and Lopez-Poveda, 2022). In a noise background, auditory neurons can shift their dynamic range towards the background sound level (Evans, 1975) and/or towards the most frequent level in the environment (Dean et al., 2005). Dynamic range adaptation can occur throughout the auditory system, from the

[3] The content of this chapter has been adapted from the following paper: López-Ramos, D., Marrufo-Pérez, M.I., Eustaquio-Martín, A., López-Bascuas, L.E. and Lopez-Poveda, E.A. (in press). Adaptation to noise in spectrotemporal modulation detection and word recognition. *Trends in Hearing*. Submitted 27 November 2023. Accepted 12 June 2024. https://doi.org/10.1177/23312165241266322

auditory nerve (AN) (Evans, 1975; Costalupes et al., 1984; Gibson et al., 1985; Wen et al., 2009, 2012), to the inferior colliculus (Rees and Palmer, 1989; Dean et al., 2005, 2008), to the auditory cortex (Phillips and Hall, 1986; Watkins and Barbour, 2008,2011). Several physiological mechanisms could be involved in dynamic range adaptation (DRA), including the activation of middle-ear muscle reflex (MEMR) (Costalupes et al., 1984; Gibson et al., 1985; Grange et al., 2022), the activation of the medial olivocochlear reflex (MOCR) (Dean et al., 2005, 2008; Wen et al., 2009, 2012; Zilany and Carney, 2010; Grange et al., 2022), or the dynamics of synaptic processes related to the exocytosis and endocytosis of neurotransmitter vesicles between cochlear inner hair cells (IHCs) and AN fibers (Dean et al., 2005, 2008; Wen et al., 2009, 2012; Zilany and Carney, 2010). Dynamic range adaptation can help encoding sound intensity (Dean et al., 2005, 2008; Wen et al., 2009, 2012), and thus TM and SM, as illustrated in **Fig. 1.1**.

Speech signals contain TM, SM, and STM. Temporal modulations pertain to variations in energy over time, while SM manifest as changes in energy over frequency. Spectrotemporal modulations combine SM and TM, thus appear in spectrograms as diagonal energy shifts (**Fig. 1.2**). Speech recognition depends primarily on the information contained in the low SM (<4 cycles/kHz) and slow TM (<12 Hz) (e.g., Elliot and Theunissen, 2009). The ability of the auditory system to accurately encode the pattern of modulations contained in speech is paramount for successful speech recognition. Indeed, recognition worsens when speech is spectrally (Ainsworth and Millar, 1972; van Veen and Houtgast, 1985; Zahorian and Jagharghi, 1993; Liu and Eddins, 2008) or temporarily degraded (Drullman et al., 1994). [Of course, it is not possible to reduce the temporal information in speech without affecting the spectral information or vice versa (Drullman et al., 1996).] Therefore, any process that improves the encoding of speech modulations, like, for example, DRA, could facilitate speech recognition.

It is not yet clear whether adaptation to noise in speech recognition results from an improvement in the encoding of speech TM, SM, or STM. Regarding temporal modulations, previous studies have shown that for AM signals in noise, AM detection improves when the signal is delayed in the noise (Marrufo-Pérez et al., 2018b) or preceded by a noise precursor (Almishaal et al., 2017; Wojtczak et al., 2019). These results are consistent with an improvement in the encoding of temporal modulations related to DRA in auditory neurons (**Fig. 1.1**). Because of this, one might think that adaptation to noise in speech recognition is mediated by an enhancement of the speech temporal modulations when speech is delayed in the noise. However, as shown in **Fig. 1.1**, DRA could also enhance the encoding of SM. Indeed, computer model simulations (Ainsworth and Meyer, 1994) show that automatic speech recognition improves as the speech spectral contrast in the AN rate profile improves when AN fibers shift their thresholds towards the noise level. However, unlike for TM, it is yet to be empirically shown that for signals in noise, the sensitivity to SM or STM improves as the signal is delayed in the noise.

One aim of the present study was to experimentally investigate if there is adaptation to noise in the detection of SM and STM. Another aim was to investigate which type of modulation cue may be involved in adaptation to noise in speech recognition. We hypothesized that if adaptation occurs for all three types of modulations (temporal, spectral, and spectrotemporal), then adaptation in word recognition might be related to improvements in the encoding of one or more of those three modulation cues. However, if adaptation does not occur for SM or STM,

then this would suggest that the modulation cue in question is unlikely involved in adaptation to noise in speech recognition.

## 4.2. MATERIAL AND METHODS

### *4.2.1. Experimental design and hypotheses*

We measured noise adaptation in word recognition and modulation detection for the same participants (paired design). Adaptation in modulation detection was measured using spectral, temporal, and spectrotemporal ripple noise. This stimulus was chosen, first, because the sensitivity to spectrotemporal ripples can partly predict speech recognition in noise (Henry et al., 2005; Bernstein et al., 2013, 2016; Mehraei et al., 2014; Davies-Venn et al., 2015; Zaar et al., 2023); and, second, because the use of ripple noise allows measuring noise adaptation in the detection of TM and SM when the two types of modulation occur separately, but also when they occur simultaneously (see below). Therefore, the use of ripple noise allows investigating whether noise adaptation differs depending on the nature of the modulation.

Adaptation in speech recognition was measured using isolated words that were unprocessed (natural) or processed (vocoded) to have poor spectral and temporal fine structure (TFS) information but preserve temporal envelope cues (Shannon et al., 1995). Adaptation was measured for these two types of words, first, because adaptation is overall greater for vocoded than for natural words (Marrufo-Pérez et al., 2018a; 2020). A second reason is that vocoded-word recognition relies mostly on detecting and processing TM, much less so on SM, and not at all on TFS speech cues (Shannon et al., 1995). Therefore, adaptation in modulation detection may be more relevant for adaptation in the recognition of vocoded than natural words.

Adaptation to noise is usually quantified as the improvement in performance when the target signal is delayed in the noise (reviewed by Marrufo-Pérez and Lopez-Poveda, 2022). The same noise adaptation paradigm was employed across all tasks to facilitate the comparison of the results in the word recognition and modulation detection tasks. Specifically, white noise (1-22.05 kHz) at 60 dB SPL was used as the background noise. Noise adaptation was quantified as the difference in performance between an 'early' and a 'late' condition. In the early condition, the target signal (word or ripple noise) was presented 50 ms after the onset of the noise. In the late condition, the onset of the target signal was delayed 800 ms from the noise onset. In both conditions, the background noise ended 50 ms after the end of the signal (**Fig. 4.1**).

### *4.2.2. Participants*

Eighteen NH adults (five men) participated in the experiments (mean ± SD age = 26.6 ± 5.6 years). Seventeen of them had audiometric thresholds ≤20 dB hearing level (HL) at octave frequencies between 125 Hz and 8 kHz in the test (left) ear (ANSI, 1996). One participant had an audiometric threshold of 25 dB HL at 8 kHz. Sixteen participants were native speakers of Spanish, while the other two were native speakers of Portuguese but were proficient in Spanish (their speech recognition and adaptation scores were not outliers according to Tukey's rule).

They were not paid for their services. All participants signed an informed consent form to participate in the study

### 4.2.3. Test sessions

The experimental protocol consisted of multiple sessions in which participants performed word recognition and/or modulation detection tasks. The first session was devoted to performing the tasks in quiet, which allowed participants to become familiar with the stimuli and tasks. Subsequent sessions were devoted to performing the tasks in noise. In these sessions, the early and late conditions were always measured in pairs but were administered in random order. Participants completed every task three times in three separate blocks. In each block, tasks were administered in random order. The total duration of the experiment was approximately six hours. Participants took breaks as needed.

### 4.2.4. Spectrotemporal modulation task

**Stimuli**

The target signal was a broadband ripple noise like the one used by Chi et al. (1999). The ripple noise consisted of 92 random-phase tones equally spaced along the logarithmic frequency axis between 140 Hz and 7.34 kHz (5.75 octaves).

The ripple profile in the temporal and spectral domains is determined by the following equation (**Eq. 4.1**):

$$S(x,t) = A \cdot \sin[2\pi(\omega t + \Omega x) + \phi] \quad (4.1)$$

where $x$ is the position on the logarithmic frequency axis (in octaves): $x=\log_2(f/f_0)$ with $f_0$=140 Hz as the lower edge of the spectrum, and $f$ as the frequency; $t$ represents time; $A$ denotes the amplitude of the ripple profile relative to a flat ripple (or modulation depth), with $A$=0 corresponding to the unmodulated ripple, and $A$=1 corresponding to a fully modulated ripple; $\omega$ is defined as the ripple velocity (number of ripple cycles per second) in Hz; $\Omega$ is the ripple density in cycles per octave (c/o); and $\phi$ represents the phase of the profile, which was set to a random value between 0 and 2π radians every time a ripple noise was generated.

The duration of the ripple noise was 200 ms, and its overall level 60 dB SPL. Therefore, the signal-to-noise ratio (SNR) in the noise tasks was 0 dB. In the SM detection task, $\Omega$ was set to 2 c/o, whereas $\omega$ was set to 0 Hz (i.e., no TM). In the TM detection task, $\Omega$ was set to 0 c/o (i.e., no SM) and $\omega$ was set to 10 Hz. In the STM detection task, $\Omega$ was set to 2 c/o and $\omega$ was set to 10 Hz, which resulted in a downward-moving ripple. **Figure 4.1** shows the spectrograms of the stimuli for each modulation detection task. Low ripple density and velocity were chosen because low SM and slow TM contribute the most to speech intelligibility (Elliot and Theunissen, 2009). Furthermore, SM sensitivity in quiet at the same ripple density of this study (2 c/o) showed a significant correlation with speech intelligibility in noise (Bernstein et al., 2013, 2016; Zaar et al., 2023). The ripple velocity (10 Hz) used here was chosen to guarantee at least two AM cycles over the used stimulus duration (200 ms). This ripple velocity was

intermediate to the values used in previous studies [4 Hz in Zaar et al. (2023) and up to 32 Hz in Bernstein et al. (2013)].

In this task, the duration of the background noise was 300 ms in the early condition (50-ms signal-noise onset delay, plus 200-ms signal duration, plus 50-ms noise-signal offset delay) and 1050 ms in the late condition (800 ms + 200 ms + 50 ms).

**Procedure**

Modulation detection thresholds were measured using a three-alternative forced-choice (3AFC) procedure. Each trial consisted of three observation intervals. The modulated ripple noise was randomly assigned to one of them, and the other two intervals contained the unmodulated ripple noise. The participant had to choose the interval in which the modulated stimulus was presented. Feedback was provided on the correctness of the response. The modulation depth ($A$) in dB full scale (FS), i.e., $L_A=20\cdot\log_{10}(A)$, was varied adaptively following a "two-down, one-up" rule (i.e., the modulation depth decreased after two correct responses and increased after an incorrect response) that leads to the 70.7% correct response point on the psychometric curve (Levitt, 1971). The initial modulation depth was set to $A$=1 (0 dB), i.e., the ripple noise was fully modulated. The procedure continued until 12 reversals in modulation depth were recorded. For the first six reversals, the modulation depth changed in 4-dB steps; for the remaining reversals, the step was 2 dB. The detection threshold was calculated as the mean modulation depth at the last eight reversals. The threshold for each participant and condition was the mean of three measurements. At the start of each trial, a pure tone (1 kHz) was presented to alert the participant as to the start of the stimulus presentation.

**Control experiment with upward-moving ripple**

While some studies have suggested that upward-moving ripples are more easily detected than downward-moving ripples (Chi et al., 1999; Zaar et al., 2023), other studies suggest that the direction of the ripple has no effect on modulation detection thresholds (Bernstein et al.,2013; Mehraei et al., 2014; Oetjen and Verhey, 2015). Here, a control experiment was performed to investigate if the direction of the ripple could affect the magnitude of adaptation to noise in STM detection. In this control experiment, Ω was set to 2 c/o and $\omega$ was set to -10 Hz (notice the negative sign), which resulted in an upward-moving ripple. Otherwise, the stimuli and procedure were as in the main experiment. Thirteen NH adults performed this control experiment, two of whom had completed the main experiment.

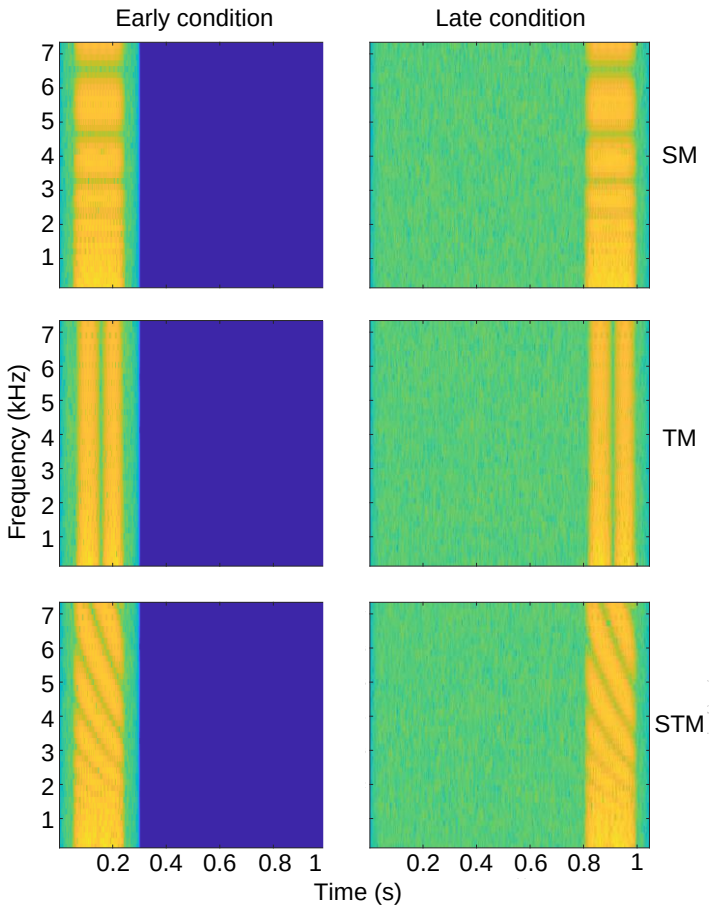


**Figure 4.1.** Example spectrograms of the stimuli used in the spectral (SM) (first row), temporal (TM) (second row), and spectrotemporal modulation (STM) detection tasks (third row). The target signal (broadband ripple noise) is shown in yellow. The background noise (white noise) is represented in green, and silence is shown in blue. The left column illustrates the stimuli in the early condition, with a 50-ms noise-signal onset delay. The right column illustrates the stimuli in the late condition, with an 800-ms noise-signal onset delay. The noise always ended 50 ms after the signal offset. For illustration purposes, the modulation phase was equal for all the stimuli.

### *4.2.1. Speech recognition task*

**Stimuli**

Speech reception thresholds (SRTs) were measured using the corpus of phonetically balanced, Castilian-Spanish disyllabic words of Cárdenas and Marrero (1994). Twenty-five words corresponding to one of the ten available lists were used to measure each SRT. The same corpus was used to measure SRTs for vocoded words. The vocoder (Shannon et al., 1995) included a high-pass pre-emphasis filter (first-order Butterworth filter with a 3-dB cutoff frequency of 1.2 kHz), a bank of 12 sixth-order Butterworth band-pass filters whose 3-dB cutoff frequencies followed a modified logarithmic distribution between 100 Hz and 8.5 kHz, and envelope extraction via full-wave rectification and low-pass filtering (fourth-order Butterworth low-pass filter with a 3-dB cutoff frequency of 400 Hz). The envelope for each frequency channel was used to modulate the amplitude of a sinusoidal carrier at the channel center

frequency. The modulated carriers were filtered again through the corresponding band-pass filter, and sample-wise added to obtain the vocoded speech.

Eighteen SRTs were measured for each participant: 2 word types (natural and vocoded) × 3 conditions (quiet, early in noise, and late in noise) × 3 measurements per word type per condition. Because only ten different word lists were available, it was necessary to repeat eight lists for each participant. Words were presented in random order across conditions to minimize the chance that participants remembered them.

**Procedure**

The SRT in noise was defined as the SNR (in dB) at 50% recognition (Levitt, 1971). It was measured by adaptively varying the speech level using a "one-down, one-up" rule. The task started with an initial SNR of 20 dB. The speech level was varied in 4-dB steps for the first 14 words, and in 2-dB steps for the last 11 words. The SRT was calculated as the mean SNR for the last 17 words. At the beginning of each trial, a 1-kHz pure tone was presented as a warning. Participants were not given feedback on the correctness of their responses. The SRT reported for each participant is the mean of three SRT estimates.

### *4.2.2. Equipment*

During the experiments, participants were seated in a double-walled sound-attenuating chamber. Stimuli were digitally stored (word recognition) or generated (modulation detection) and presented using custom-made MATLAB software (R2017b, The MathWorks). Stimuli were played via an RME Fireface UCX soundcard at a sampling rate of 44.1 kHz with a 24-bit resolution. Stimuli were presented monaurally to the participants' left ear through Sennheiser HD-580 headphones. The equipment was calibrated by placing the headphones on a KEMAR (Knowles Electronics) equipped with a Zwislocki DB-100 coupler (Knowles Electronics) connected to a sound level meter (Brüel Kjaer, mod. 2238). Calibration was performed at 1 kHz, and the obtained sensitivity was used for all other frequencies.

### *4.2.3. Statistical analysis*

Statistical analyses were performed using the *rstatix* (Kassambara, 2021) package in R (version 4.1.0). Paired Student's *t* tests with Bonferroni correction for multiple comparisons were used for pairwise comparisons of thresholds in quiet. Since SRTs were expected to be worse for vocoded than for natural words (Marrufo-Pérez et al., 2018a; 2020), one-tailed *t* tests were used to compare SRTs for natural and vocoded words. Otherwise, *t* tests were two-tailed.

To assess the statistical significance of noise adaptation in the different tasks, a two-way repeated measures analysis of variance (RMANOVA) was performed, having the type of task (SM detection, TM detection, STM detection, natural word recognition, vocoded word recognition) and the noise-signal onset delay (early and late) as factors. When the sphericity condition was not met, the Greenhouse-Geisser correction was applied. To test for the effect of ripple direction, a mixed two-way ANOVA was performed, including the ripple direction (downward, upward) as a between-subjects factor and the temporal position (early, late) as a repeated measures factor.

Paired Student's *t* tests with Bonferroni corrections for multiple comparisons were applied for post hoc pairwise comparisons. The test to compare the magnitude of adaptation between natural and vocoded words was one-tailed, as we expected adaptation to be greater for vocoded words (Marrufo-Pérez et al., 2018a; 2020). All other post hoc tests were two-tailed.

## 4.3. RESULTS

### *4.3.1. Thresholds in quiet*

**Figure 4.2A** shows the SRTs for natural and vocoded words in quiet. The mean SRTs were significantly lower (better) for natural (mean ± SD = 22.7 ± 2.9 dB SPL) than for vocoded words (27.1 ± 4.0 dB SPL) (paired Student's *t-test*: $t_{(17)}$= -8.08, $p$<0.001).

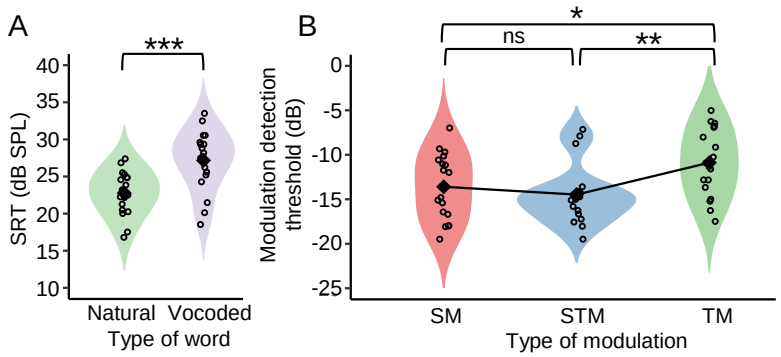


**Figure 4.2.** Violin plots showing the distribution of speech reception thresholds (SRTs) for natural and vocoded words (**A**) and detection thresholds for spectral (SM), temporal (TM), and spectrotemporal modulations (STM) (**B**) in quiet condition. Each violin plot shows the distribution of thresholds (in a distinct color for each stimulus) as well as the individual data (empty circles) and the mean (filled black diamonds). * p≤0.05; ** p≤0.01; *** p≤0.001; ns= not significant.

**Figure 4.2B** shows modulation detection thresholds in quiet. The mean TM detection threshold (-10.8 ± 3.7 dB) differed significantly from both the mean SM detection threshold (-13.6 ± 3.6 dB) (paired Student's *t-test*: $t_{(17)}$= -3.38, $p$=0.012) and STM detection threshold (-14.5 ± 3.3 dB) (paired Student's *t-test*: $t_{(17)}$= -4.25, $p$=0.002). However, the mean SM and STM detection thresholds in quiet were not significantly different from each other (paired Student's *t-test*: $t_{(17)}$= 1.11, $p$=0.852).

### *4.3.2. Adaptation to noise*

**Figure 4.3A** shows the SRTs in the early and late conditions for natural and vocoded words, while **Fig. 4.3B** shows thresholds in the early and late conditions for SM, TM, and STM detection. A two-way RMANOVA with temporal position and task as factors revealed that mean thresholds were overall lower (better) in the late condition than in the early condition ($F_{(1,17)}$=91.13, $p$<0.001) (i.e., a significant effect of adaptation to noise). In addition, the interaction between task and temporal position was also significant ($F_{(4,68)}$=7.20, $p$<0.001), which indicated that the magnitude of adaptation to noise was different across tasks. *Post hoc*

comparisons showed that the SRT improvement in the late condition was not significant for natural words (paired Student's *t-test*: $t_{(17)}$=1.90, $p$=0.19), but was significant for vocoded words (paired Student's *t-test*: $t_{(17)}$=5.44, $p$<0.001). Adaptation to noise was statistically significant for SM (paired Student's *t-test*: $t_{(17)}$= 5.19, $p$<0.001) and TM (paired Student's *t-test*: $t_{(17)}$=4.37, $p$=0.002) detection, but not for STM detection (paired Student's *t-test*: $t_{(17)}$=-0.19, $p$=1.0).

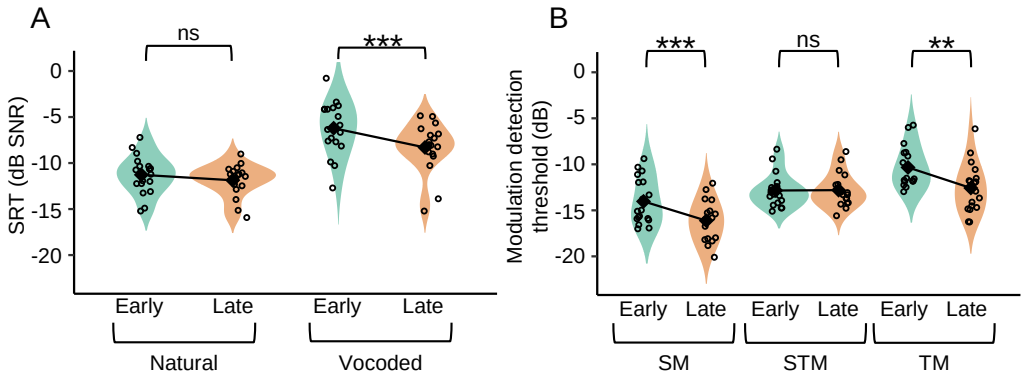


**Figure 4.3.** Violin plots showing the distribution of speech reception thresholds (SRTs) for natural and vocoded words (**A**) and detection thresholds for spectral (SM), temporal (TM), and spectrotemporal modulations (STM) (**B**) in the early (green) and late (orange) condition. Each violin plot shows the distribution of thresholds as well as the individual data (empty circles) and the mean (filled black diamonds). ** p≤0.01; *** p≤0.001; ns= not significant.

**Figure 4.4** shows the magnitude of adaptation to noise in the word recognition tasks (**A**) and modulation detection tasks (**B**). Mean adaptation was significantly less for natural (0.59 ± 1.3 dB) than for vocoded words (2.1 ± 1.6 dB) (paired Student's *t-test*: $t_{(17)}$=-3.50, $p$=0.004). This is consistent with previous studies that also found greater adaptation to noise in vocoded word recognition than in natural word recognition (Marrufo-Pérez et al., 2018a; 2020). Regarding modulation detection, noise adaptation in SM (2.1 ± 1.7 dB) and TM (2.2 ± 2.1 dB) detection was significantly different from adaptation in STM detection (-0.04 ± 0.9 dB) (paired Student's *t-test*: $t_{(17)}$=4.56, $p$=0.001; paired Student's *t-test*: $t_{(17)}$=4.15, $p$=0.003, respectively). The difference between adaptation to noise in SM and TM detection was not significant (paired Student's *t-test*: $t_{(17)}$=-0.21, $p$=1.0).

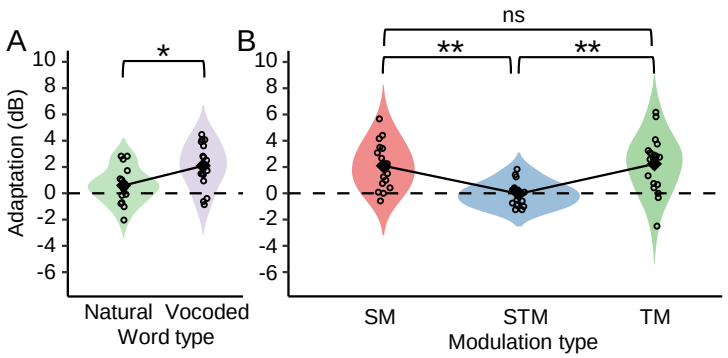


**Figure 4.4.** Violin plots showing the distribution of the values of adaptation in natural and vocoded word recognition (**A**) and in spectral (SM), temporal (TM), and spectrotemporal (STM) detection (**B**). Each violin plot shows the distribution of the values (in a distinct color for each stimulus) as well as the individual data (empty circles8 and the mean (filled black diamonds). * p≤0.05; ** p≤0.01; ns= not significant.

**Control experiment with upward-moving ripple**

**Figure 4.5A** shows the modulation detection thresholds in the early and late conditions for the upward-moving STM ripple noise. Corresponding thresholds for the downward-moving ripples were shown in **Fig. 4.3B** (STM data set). A mixed, two-way ANOVA with temporal position as repeated-measures factor and direction (downward or upward) as inter-subject factor revealed non-significant effects of temporal position ($F_{(1,29)}$=1.51, $p$=0.229) or ripple direction ($F_{(1,29)}$=1.70, $p$=0.203) as well as a non-significant interaction of these two factors ($F_{(1,29)}$=1, $p$=0.327). Adaptation measured with upward-moving ripple (-0.4 ± 1.1 dB) was not statistically significant (paired Student's *t-test*: $t_{(17)}$=-1.32, $p$=0.894) and did not differ from the one measured with downward-moving ripple (Student's *t-test*: $t_{(29)}$= 1, $p$= 0.327) (**Fig. 4.5B**). In addition, modulation detection thresholds in quiet (data not shown) did not differ for downward- and upward-moving ripples (Student's *t-test*: $t_{(29)}$=-0.253, $p$=0.802), consistent with results reported in previous studies (Bernstein et al.,2013; Mehraei et al., 2014; Oetjen and Verhey, 2015). Thus, we conclude that the lack of an improvement in STM detection when delaying the signal from the noise onset does not depend on the direction of the modulation.

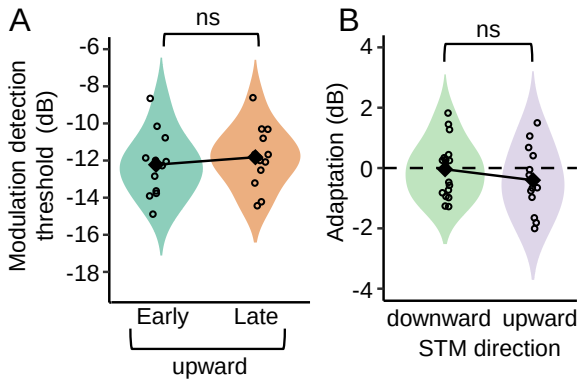


**Figure 4.5.** Violin plots showing the distribution of detection thresholds for upward spectrotemporal modulations (STM) (**A**). The distribution of the values of adaptation in downward (replotted from **Figure 4.4B**, STM dataset) and upward STM detection (**B**). Each violin plot shows the distribution of the values as well as the individual data (empty circles) and the average (filled black diamond). ns= not significant.

## 4.4. DISCUSSION

One aim of the present study was to investigate if there is adaptation to noise in the detection of TM, SM, and STM using ripple noise. We hypothesized that if neural DRA participates in adaptation to noise, it should improve the neural representation, thus the detection of SM as well as TM (**Fig. 1.1**). We found that noise adaptation occurs in the detection of TM and SM, but not in the detection of combined STM (**Fig. 4.3B**).

Another aim was to investigate which type of modulation may be involved in adaptation to noise in word recognition. We hypothesized that word-in-noise recognition may be better in the late than in the early condition because DRA improves the neural encoding of speech (SM, TM, and/or STM) modulation cues in the late condition. If this were the case, adaptation to noise in modulation detection should occur for all three types of modulations. Consistent with previous studies, we found statistically significant adaptation in the recognition of vocoded words, but not in the recognition of natural words (**Fig. 4.3A**). However, in contrast to expectations, for the same listeners, adaptation to noise occurred in SM and TM detection but not in STM detection (**Fig. 4.3B**). This suggests that STM detection cues may not be involved in adaptation to noise in speech recognition.

### *4.4.1.Adaptation to noise in modulation detection*

As far as we know, this is the first study that empirically demonstrates an improvement in SM detection as the SM signal is delayed from the noise onset. Our findings, however, are consistent with the computer model simulations of Ainsworth and Meyer (1994). They showed that the speech spectrum was better encoded in the AN rate profile when speech was presented in continuous than in gated background noise. In their model, the improvement was the result of including a shift in the threshold and dynamic range of AN fibers towards the noise level in the continuous background noise condition.

Regarding TM detection, the mechanism for detecting temporal ripples superimposed on a noise could be the same as that used for detecting sinusoidal AM superimposed on a tone. Unfortunately, a direct comparison between noise adaptation in TM detection as measured here with the adaptation found in previous AM detection studies is not straightforward because the amount of adaptation depends on the modulation rate (Viemeister, 1979; Sheft and Yost, 1990; Almishaal et al., 2017), the sound level (Almishaal et al., 2017; although see Wojtczak et al., 2019), the laterality and type of noise (Marrufo-Pérez et al, 2018b; Wojtczak et al., 2019), the stimulus type [pure tone (Marrufo-Pérez et al, 2018b; Wojtczak et al., 2019) vs. noise (Viemeister, 1979; Sheft and Yost, 1990; Almishaal et al., 2017)], and the testing paradigm [use of precursors (Almishaal et al., 2017; Wojtczak et al., 2019) vs. delay from noise onset (Marrufo-Pérez et al, 2018b)]. For a review of adaptation to noise in AM detection, see Marrufo-Pérez and Lopez-Poveda (2022).

The improvement in sensitivity to SM and TM as they are delayed from the noise onset could be explained by a noise-induced shift of the neural dynamic range (as shown in **Fig. 1.1**). But why would DRA not enhance the sensitivity to combined STM? The reason is unknown, but **Fig. 4.6** illustrates a hypothetical explanation based on DRA in the AN (Ainsworth and Meyer, 1994; Marrufo-Pérez et al., 2018b). The figure schematically illustrates the AN representation of SM, TM, and STM stimuli presented early and late in noise. The shown rate-level functions are for a hypothetical AN fiber with (dotted lines) and without (continuous lines) a noise-induced shift of the threshold and dynamic range towards the noise level ($N_L$). As the stimuli are "passed" through rate-level functions for the population of AN fibers, DRA would enhance the three kinds of modulations as illustrated by the greater contrast between peaks and valleys in the AN representation in the late than in the early condition. There exist central auditory neurons that integrate neural responses across time (e.g., multipolar neurons in the cochlear nucleus, Blackburn and Sachs, 1990; or neurons in the central nucleus of the inferior colliculus, Rodríguez et al., 2010) or characteristic frequency (CF) (e.g., octopus cells in the cochlear nucleus, Oertel et al., 2000; or neurons in the central nucleus of the inferior colliculus, Rodríguez et al., 2010) presumably to provide stable representations of the stimulus spectrum and envelope. If modulation sensitivity depends on integrating neural population responses, a greater sensitivity to SM is predicted in the late condition when the AN responses are integrated across time. Likewise, a greater sensitivity to TM in the late condition is predicted when the AN responses are integrated across CFs. However, if STMs are detected by the same mechanism as SMs (i.e., by resolving spectral peaks and valleys) (Summers and Leek, 1994; Eddins and Bero, 2007; Bernstein et al., 2013,2016; Davies-Venn et al., 2015; Narne et al., 2016, 2018; Isarangura et al., 2019) little or no change in sensitivity to STM is predicted in the late condition. The same would happen if STMs were detected by using the integrated AN response across CFs.

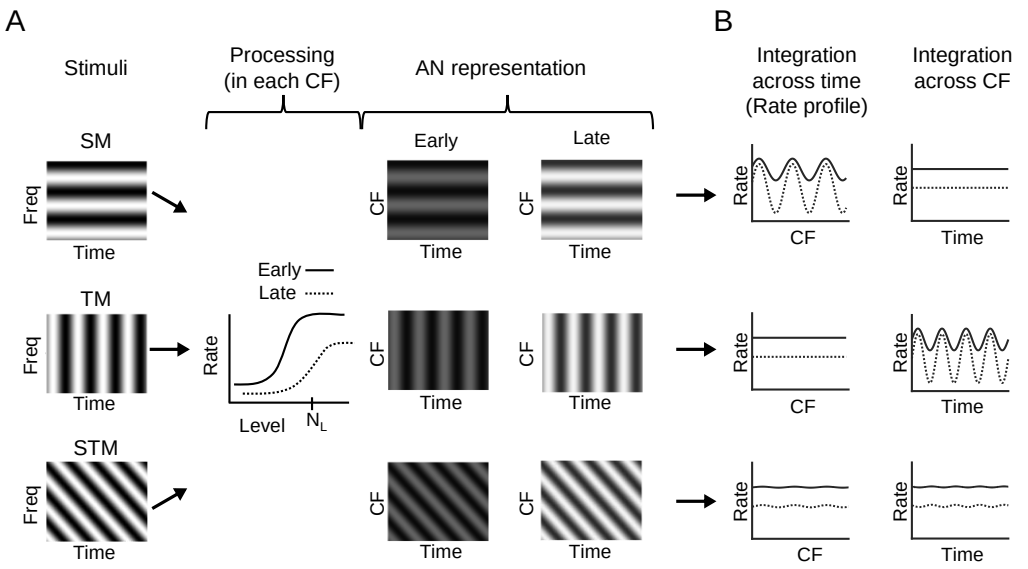


**Figure 4.6.** Hypothetical explanation for the absence of adaptation to noise in the detection of spectrotemporal modulation (STM). Schematic illustration of the effect of a shift in the dynamic range of auditory nerve (AN) fibers on the AN representation of spectral (SM) (first row), temporal (TM) (second row) and STM (third row). Rate-level functions are shown with (dotted lines) and without (continuous lines) a hypothetical noise-induced shift of their threshold and dynamic range. The AN representation of the three modulations is enhanced (higher contrast) in the late condition than in the early condition (lower contrast) because of this shift in the dynamic range of the fibers (**A**). Schematic representation of each modulation when AN spikes are integrated across time (rate profile, first column) and across CFs (second column). After integration, spectral and temporal envelopes are better represented in the late condition for spectral and temporal modulations but not for spectrotemporal modulations (**B**). See the main text for details.

The detection of SM and STM may not always depend on the ability to resolve the spectral peaks and valleys of the modulation. Several cues such as local spectral loudness, a different spectral centroid between the reference and target stimuli, and/or differences at the spectral boundaries of the reference and target stimuli (Azadpour and Mckay, 2012; Aronoff and Landsberger, 2013) can influence the discrimination/detection of SM. In addition, STM detection could be mediated by an AM cue at the output of the auditory filters at the lower and upper spectral limits of the ripple noise (Narne et al., 2016). In the present study, local loudness cues, or differences in the spectral centroid or at the spectral boundaries were unreliable cues for SM detection because a random phase was applied each time the ripple noise was generated (Aronoff and Landsberger, 2013). Likewise, the AM cue at the output of the auditory filters at the spectral edges of the ripple noise would be masked by the noise. Hence, here, STM detection probably relied on resolving spectral peaks, i.e., on the same mechanism proposed to detect SM. In fact, because the background noise was white and the ripple noise had a downward-sloping spectra, SM and STM detection probably involved resolving the spectral peaks in the low-frequency region, where the signal-to-noise ratio was higher.

Of course, the hypothetical mechanism illustrated in **Fig.4. 6** is an oversimplification and almost certainly inaccurate. For example, as it stands, it would predict worse thresholds for STM detection than for SM or TM detection because spectrotemporal modulations would be

smeared when integrating neural responses across time and/or CF. This prediction is not consistent with the experimental data, as STM detection thresholds were close to SM detection thresholds and better than TM detection thresholds. However, **Fig. 4.6** represents a hypothetical extreme case where the spectral and temporal integration windows span the entire stimulus. It is likely that the auditory system integrates neural responses using smaller windows, thus the loss of STM information is probably less than suggested by **Fig. 4.6**. In addition, Bernstein et al. (2013) found that the sensitivity to frequency modulation (FM) at 2 Hz applied to a 500-Hz pure-tone carrier was a significant predictor of STM sensitivity. Assuming that FM sensitivity at low modulation rates is encoded in the timing of neural spikes via phase locking (e.g., Moore and Sek, 1996; but see Whiteford et al., 2020), the findings of Bernstein et al. (2013) suggest that STM sensitivity might depend on the ability to use TFS cues. This possibility is disregarded in **Fig. 4.6**.

Other phenomena besides neural DRA may have contributed to the current results. For example, thresholds might have been better in the late than in the early condition because the background noise functioned as warning and drew the listener's attention to the task. However, we tried to minimize this effect by presenting a pure tone cue at the beginning of each trial. Furthermore, if this attentional driven improvement in sensitivity had occurred, it would have affected performance equally in all modulation detection tasks. The fact that detectability in the late condition improved for SM and TM detection but not for STM detection undermines this mechanism as an explanation for adaptation to noise. Modulation masking could have also contributed to the current results, i.e., the onset of background noise could be interpreted as a temporal modulation (Gallun and Hafter, 2006). The onset of the background noise was closer in time to the signal in the early than in the late condition. This could have caused some modulation masking in the early condition, making it seem that modulation sensitivity was worse in the early than in the late condition. However, while this phenomenon could have affected TM detection, it is unclear how it could have affected SM or STM detection.

### *4.4.2. Less noise adaptation for natural than for vocoded words*

Like previous studies (e.g., Marrufo-Pérez et al., 2018a, 2020), we have found greater adaptation to noise in vocoded than in natural word recognition. Two explanations have been proposed elsewhere for this result. The first one (proposed by Marrufo-Pérez et al., 2018a, 2020) relates to a ceiling effect: because SRTs tend to be better (lower) for natural than vocoded words, there is less room for SRT improvement for natural than vocoded words. The second explanation (proposed by Marrufo-Pérez et al., 2020) is that improved envelope cues may be less effective when word recognition relies in part on TFS encoding, as may be the case for natural words (Hopkins and Moore, 2009).

Intriguingly, the present data show that explanation #1 does not hold for modulation detection tasks as thresholds in the early condition were nearly identical for SM and STM detection, and adaptation was statistically significant only in SM detection. Moreover, although thresholds were worse for TM than SM detection in the early condition, adaptation was similar for the two tasks. Explanation #2, however, would be supported by the present data if TFS information played a role in STM detection (Bernstein et al., 2013), as we found no adaptation to noise in STM detection or in the recognition of natural words.

### *4.4.3. Relationship between adaptation in modulation detection and speech recognition*

Dynamic range adaptation can improve the neural encoding of SM and TM (**Fig. 1.1**). Both SM and TM are important for speech recognition. Therefore, if DRA is involved in adaptation to noise, one would expect adaptation in modulation detection to occur for the same listeners who show adaptation in speech recognition. We found that noise adaptation occurs for SM and TM detection as well as for vocoded word recognition. This indicates that noise adaptation in (vocoded) word recognition may be mediated by improvements in the neural encoding of sustained SM and/or TM speech cues. However, adaptation did not occur in STM detection (**Fig. 4.4B**). This indicates that it is unlikely that adaptation to noise in (vocoded) word recognition be mediated by improvements in the neural encoding of STM speech cues.

The present data and design were insufficient to elucidate which of the two modulation cues (SM or TM) for which adaptation occurred, if any, is more prominently involved in adaptation in (vocoded) word recognition. This could be investigated, for example, by quantifying the correlation between adaptation in modulation detection and word recognition. Based on the present data, adaptation word recognition was not correlated with adaptation in modulation detection (**Fig. 4.7**). However, a larger participant sample or a sample with more variable adaptation scores would be needed to address this question using correlational analyses with enough statistical power.

There are other factors that could explain the lack of correlation between adaptation in both tasks: First, different methods were employed to assess word recognition and modulation detection. For example, the SNR was fixed at 0 dB in the modulation detection task while it varied in the word recognition task to measure the individual SRT (the group mean SRT was about -10 dB). Second, the present modulation detection tasks assess a listener's sensitivity to sustained modulations. However, transient information, such as vowel formant transitions, is essential for speech perception. For example, Varnet et al. (2013) showed that listeners use mainly the end of the second formant (F2) in the first vowel and the onset of F2 in the second vowel as cues to discriminate between /aba/ and /ada/. Liberman et al. (1954) also showed that F2 transitions are key for distinguishing between these two phonemes. In other words, the recognition of disyllabic words in noise could depend on detecting a word's transient features as well as on detecting their (long-term) temporal or spectral envelopes. Therefore, adaptation to noise in word recognition may also be related to improvements in the neural encoding of speech transients, something that was not explored here.

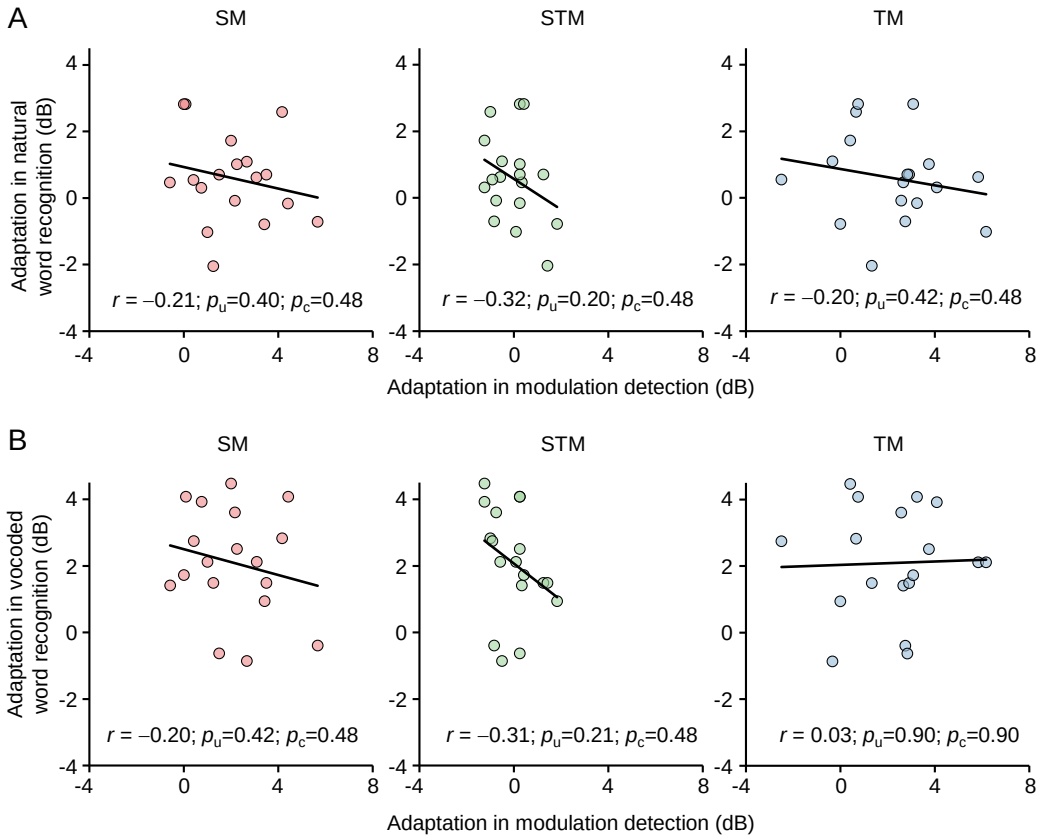


**Figure 4.7.** Relationship between adaptation to noise in modulation detection and speech recognition for natural (**A**) and vocoded words (**B**). N=18. Insets give the Pearson correlation coefficient ($r$) and corresponding significance levels uncorrected ($p_u$) or corrected ($p_c$) for multiple comparisons controlling the false discovery rate (Benjamini and Hochberg, 1995). SM, Spectral Modulation; STM, Spectrotemporal Modulation; TM, Temporal modulation.

## 4.5. CONCLUSIONS

1) For signals in noise, the sensitivity to SM and TM but not to STM improves as the signal is delayed 800 ms from the noise onset. In other words, adaptation to noise occurs in the detection of SM and TM but not in the detection of combined STM.
2) The recognition of words in noise can improve as words are delayed 800 ms in the noise. The improvement was statistically significant for vocoded but not for natural words.
3) Therefore, adaptation to noise in word recognition may be mediated by improvements in the neural encoding of sustained SM or TM, but it is unlikely mediated by improvements in the encoding of combined STM.

# 5.

# GENERAL DISCUSSION

In noise backgrounds, auditory neurons can shift their dynamic ranges towards the level of the noise and towards the most common level in the environment (e.g., Evans, 1975; Phillips and Hall, 1986; Dean et al., 2005). Several physiological mechanisms, including the middle-ear muscle reflex (MEMR), the medial olivocochlear reflex (MOCR), and/or synaptic processes, may participate in dynamic range adaptation (DRA). DRA can prevent the neuron's firing rate from becoming saturated and thus improve the encoding of the acoustic temporal and spectral envelopes (**Fig. 1.1**). However, it is often assumed that the ability of the auditory system to encode and process spectral or temporal information is static and unaffected by adaptation. One aim of this thesis was to empirically explore if auditory frequency selectivity and the sensitivity to spectral (SM), temporal (TM), and spectrotemporal modulations (STM) are affected by noise adaptation. Another aim was to investigate if adaptation to noise in speech recognition (i.e., the improvement in speech recognition as the speech is delayed in the noise) relates to an enhancement in the encoding of the speech temporal and/or spectral envelopes, as proposed by previous studies (Ainsworth and Meyer, 1994; Marrufo-Pérez and Lopez-Poveda, 2022).

Two studies have been conducted to investigate the possible effect of adaptation processes on human auditory frequency selectivity. The first study (described in **Chapter 2**) explored whether the stimulus duration affects estimates of human cochlear tuning at low and high frequencies. Cochlear tuning was estimated using the behavioral method of Shera et al. (2002), which involves using forward notched-noise (NN) maskers. The hypothesis was that tuning estimates may be different when the maskers are long enough to activate the MEMR and/or the MOCR. It has been shown that masker duration has a negligible effect on tuning at 4 kHz. In contrast, the use of short masker produces broader tuning estimates at 500 Hz.

The second study (described in **Chapter 3**) investigated the effect of noise precursors on human psychoacoustical tuning curves (PTCs) at 500 Hz and 4 kHz. The stimuli used to measure the PTCs were too short to activate the MEMR or the MOCR by themselves. The noise precursors, however, were designed to activate the MOCR with minimal activation of MEMR. In other words, it was assumed that, by design, any effect of the precursors on the PTCs would be due to the MOCR. It has been shown that contralateral precursors have a negligible effect on the PTCs. However, ipsilateral and bilateral precursors sharpen the PTCs at 500 Hz but broaden the PTCs at 4 kHz. That is, ipsilateral and bilateral noise precursors affect the PTCs differently at low and high frequencies. A physiologically inspired computer model of forward masking with efferent control (from Aguilar et al., 2013) was used to simulate and interpret the results. The simulated PTCs were broadly consistent with the experimental PTCs and suggested that, while

seemingly different, the pattern of results at the two test frequencies is consistent with the precursors inhibiting the gain of the cochlear amplifier by activation of the MOCR.

Lastly, an experimental study has been conducted to investigate if the sensitivity to SM, TM and STM improves as the modulated signal is delayed in the noise (**Chapter 4**). It has been shown that this is indeed the case for SM and TM detection, but not for STM detection. Furthermore, it has been shown that for the same listeners, adaptation to noise occurs in the recognition of vocoded, but not natural words. Combined, these findings suggest that adaptation to noise in speech recognition is unlikely mediated by improvements in the encoding of STM.

## 5.1. IMPLICATIONS

The finding that the behavioral estimates of human cochlear tuning can vary depending on the stimulus duration, at least at 500 Hz, has implications for both understanding human auditory frequency selectivity and designing optimal methods and stimuli to estimate it. The employed forward masking NN method (by Shera et al., 2002) is a well-controlled behavioral technique for inferring cochlear frequency tuning. Previous studies have shown that it yields tuning estimates that broadly match those inferred from the latency of stimulus-frequency otoacoustic emissions (SFOAEs) (in humans: Shera et al., 2002), as well as direct measurements from auditory nerve (AN) responses (in ferret: Sumner et al., 2018). However, the evidence for a correspondence between behavioral, SFOAEs, and neural estimates of tuning is restricted to probe frequencies ≥1 kHz and masker durations ≥ 200 ms. It remains to be shown that behavioral estimates match SFOAE or neural data at lower frequencies (<1 kHz) and for other (shorter) masker durations.

Here, we have shown that at 4 kHz the method of Shera et al. (2002) yields equal tuning estimates using short (30 ms) and long (400 ms) maskers. This indicates that at 4 kHz cochlear tuning itself or the estimates obtained with this method at 4 kHz are unlikely affected by the MOCR, the MEMR or other potentially confounding factors discussed in **Chapter 2**. In contrast, at 500 Hz tuning estimates are significantly broader using the shorter maskers. As discussed in **Chapter 2**, this dependency on masker duration could reflect a nonlinear interaction of physiological mechanisms that can affect cochlear tuning (e.g., the MEMR and/or MOCR) with acoustic (e.g., spectral splatter) or cognitive factors that may affect human performance in the masking task. It may be hard to control for these factors, as the measures that can minimize the impact of one factor might increase the impact of another. For example, using longer maskers to minimize spectral splatter would increase the chance that the estimates be affected by the MEMR or the MOCR. In other words, at low frequencies, the stimulus used to estimate tuning might itself change the tuning and/or affect the measurement. Therefore, estimates of human cochlear tuning at low frequencies (e.g., Leschke et al., 2022) should be interpreted with caution.

The demonstration that frequency selectivity and SM and TM sensitivity change in background noise (**Chapter 3** and **4**) has implications for understanding human auditory spectrotemporal processing in noise. For example, it suggests that to accurately simulate auditory spectrotemporal processing in real-world noisy environments it becomes critical to

incorporate noise adaptation mechanisms in models of the auditory periphery. Traditionally, few models have included mechanisms like the MEMR or MOCR to account for these noise-induced changes in spectrotemporal processing (see Lopez-Poveda, 2005, for a comprehensive review of models of the auditory periphery). The model used in **Chapter 3** incorporates an attenuation parameter to account for the inhibition of cochlear gain caused by activation of the MOCR. However, the gain inhibition in that model is not dynamic as the value of the efferent attenuation parameter was fixed during the entire stimulus presentation. In a more recent version of the model (Grange et al., 2022) MOCR efferent attenuation is dynamically dependent on the model response, which is physiologically more realistic. Another peripheral model that incorporates adaptation (Zilany and Carney, 2010) has successfully simulated DRA to the most frequent environmental sound level in AN neurons (Wen et al., 2009). This model has also been updated to incorporate dynamic MOCR control of cochlear gain dependent on excitatory inputs from the inferior colliculus and cochlear nucleus (Farhadi et al., 2023)

## 5.2. LIMITATIONS AND FUTURE WORK

### *5.2.1. Effect of MEMR and MOCR on frequency selectivity estimates*

The present work was aimed to investigate noise-induced changes in human auditory frequency selectivity. Specifically, we have compared frequency selectivity estimates obtained using long and short NN maskers (**Chapter 2**), and with and without noise precursors (**Chapter 3**). The observed differences have been discussed *assuming* that the longer maskers and the precursors may activate the MEMR or the MOCR. For future studies, it would be worth assessing the activation of the two reflexes directly using physiological methods rather than making assumptions about whether the used stimuli activated those reflexes. For example, the degree of MOCR activation produced by the precursors or the long-duration maskers could be quantified by measuring the suppression of otoacoustic emissions (Marrufo-Pérez et al., 2021) and the degree of activation of the MEMR could be assessed by performing a MEMR contraction test (Lilaonitkul and Guinan, 2009a,b; Aguilar et al., 2013). Those physiological measurements would allow more direct and precise assessment of the potential influence of the MEMR or MOCR on frequency selectivity estimates. They would also serve to ensure that the stimuli used to measure frequency selectivity do not activate the reflexes by themselves while the stimuli designed to activate the reflexes indeed do so.

### *5.2.2. Relationship between frequency selectivity, modulation detection, and speech-in-noise recognition*

As reviewed in **Chapter 1**, previous studies have shown a strong relationship between frequency selectivity, STM sensitivity, and speech-in-noise recognition (Bernstein et al., 2013; Davies-Venn et al., 2015). Here, only the relationship between STM sensitivity and speech recognition in noise has been studied (**Chapter 4**). Although some participants took part in all frequency selectivity, STM sensitivity, and word recognition experiments, the number of those participants was insufficient to calculate pairwise correlations between the three measures with sufficient statistical power. In addition, noise-induced changes in frequency selectivity

have not been measured using the same paradigm as the noise-induced change in modulation detection and word recognition (use of precursors vs. signal delay from noise onset). This difference in methodology precludes drawing conclusions about the relationship between these three aspects of auditory perception in noise. In the future, it would be interesting to go beyond the well-established relationship between frequency selectivity, modulation detection and word recognition in noise, and investigate whether noise-induced changes in auditory filter bandwidth are related with the noise-induced improvements in SM sensitivity and word recognition. One hypothesis would be that, if noise adaptation mechanisms improve frequency selectivity in noise, this would be related to improvements in SM sensitivity and word-recognition.

### *5.2.3. Use of white noise and broadband noise with flat spectrum*

In the present experiments, the adapting noise was broadband with a flat spectrum (**Chapter 3**) or white (**Chapter 4**). It is not clear that broadband, flat-spectrum noises are the most adequate to study adaptation to noise. On the one hand, SFOAE studies have found that broadband noises are the most efficient MOCR elicitors (Lilaonitkul and Guinan, 2009a). Also, the MEMR activation threshold tends to be lower for broadband noise elicitors (Keefe et al., 2010). On the other hand, in the present modulation detection and speech recognition tasks, the modulated noise and the speech had a downward-sloping spectra and the use of flat-spectrum noises could have forced participants to rely on low-frequency information to detect the modulation or to recognize speech. The magnitude of adaptation to noise in speech recognition reported in **Chapter 4** was close to that reported in previous studies in which the spectrum of the noise produces the same amount of masking in all the frequencies components of speech (e.g. speech-shaped noise) (Marrufo-Pérez et al., 2018a; 2020). But in the case of modulation detection, it is not clear to what extent adaptation depends on the availability of low-frequency cues. For that reason, future experiments could explore adaptation to noise in the detection of TM or SM superimposed on broadband signals using background noise with the same spectrum as the signal. It also would be interesting to explore the effect of more ecological background noises on adaptation to noise in speech recognition and modulation detection to explore the possible influence of adaptation to noise when listening in more ecological environments (Marrufo-Pérez et al., in preparation).

### *5.2.4. Adaptation to noise in SM and TM detection*

In **Chapter 4**, we examined adaptation to noise in SM and TM detection for only one combination of spectral density and temporal rate (2 c/o and 10 Hz, respectively). It would be interesting to explore in future research how adaptation to noise in both TM and SM varies depending on these parameters. As discussed in **Chapter 1,** TM and SM sensitivity changes with modulation rate and density, respectively. Specifically, the TM transfer function shows a low-pass trend where TM sensitivity remains constant between 2 and 16 Hz and gradually decrease for higher modulations rates (Viemeister, 1979), while SM transfer function peaks at 1-to-4 c/o (Summers and Leek, 1994; Eddins and Bero, 2007; Isarangura et al., 2019). Prior research indicates a great improvement in modulation detection when the modulated signal is superimposed at the end of continuous band of noise rather than gated with the noise for

modulation frequencies ≤ 20 Hz (Viemeister, 1979; Sheft and Yost, 1990). Similarly, the improvement in amplitude modulation (AM) detection when preceded by a precursor decreases as the AM rate increases from 20 Hz to 500 Hz (Almishaal et al., 2017). This gives support to the hypothesis that adaptation to noise may be directly related to sensitivity, i.e., the greater the sensitivity to AM, the greater the amount of adaptation to noise in AM detection. Although this idea contradicts previous studies, which found an inverse relationship between adaptation to noise and performance in word recognition (Marrufo-Pérez et al., 2018a, 2020), it would be worth investigating whether the relationship holds true for noise adaptation in the detection of SM and TM superimposed on broadband signals.

Another limitation of the present work is that adaptation to noise in modulation detection was measured for a fixed noise level (60 dB SPL) and a fixed signal-to-noise ratio (SNR) (0 dB). It would be worth studying if adaptation to noise in SM and TM detection varies with sound level and/or SNR. Previous studies have investigated the effect of sound level (Wojtczak et al. 2019) and SNR (Almishaal et al., 2017; Mesik and Wojtczak, 2020) on adaptation to noise in AM detection. However, it is yet to be investigated whether sound level and SNR affect adaptation when TM and SM are superimposed on broadband stimuli. These data could shed light on the contributions of different physiological mechanisms to behavioral adaptation (Marrufo-Pérez and Lopez-Poveda, 2022). For instance, if noise adaptation were related to the MEMR adaptation should be greatest at high noise levels because the MEMR is stronger at high levels. If noise adaptation were related to the MOCR then adaptation should be greater at mid-levels where cochlear compression occurs and the MOCR has a strong effect. On the other hand, if adaptation to noise were more related to neural DRA towards the noise level, then behavioral adaptation should be greater at a SNR close to 0 (Mesik and Wojtczak, 2020).

### *5.2.5. Lack of adaptation to noise in STM detection*

It has been shown in **Chapter 4** that adaptation to noise does not occur for STM detection. However, the finding was restricted to an SM density of 4 c/o and a TM rate of 10 Hz. It could be interesting to investigate whether the lack of adaptation in STM detection is limited to these parameters and whether adaptation may occur for STMs with other ratios between modulation rate and density, i.e., with other slopes in the spectrogram (see, for example, Oetjen and Verhey, 2015). It would be reasonable to expect adaptation in STM detection to occur gradually as the STM becomes more vertical (thus more like TM) or more horizontal (thus more like SM). This could shed light on the hypothesis (proposed in **Chapter 4)** that the lack of adaptation to noise in STM detection might be due to spectral and temporal integration processes in central auditory centers. These integration processes would counteract the benefit of DRA for encoding STMs. As the slope of the STM becomes steeper or flatter in the spectrogram, the detrimental effect of spectral and temporal integration would be smaller, and adaptation to noise would be greater.

The lack of adaptation to noise in STM detection may also indicate that the mechanism for detecting STM is more closely related to the mechanism for detecting frequency modulation (FM) than to the mechanism for detecting SM or TM. Bernstein et al. (2013) found that FM sensitivity to 2-Hz FM applied to a 500 Hz signal was a significant predictor of STM sensitivity. This suggests that FM and STM could be detected by similar mechanisms and thus indicate that

the use of phase locking to track changes in spectral peaks may be key for detecting STMs. While it has been shown that DRA improves the encoding of sound level in the AN (Evans, 1975; Gibson et al., 1985), it is not clear whether there exists an adaptation mechanism that could improve phase locking in the AN. If FM and STM were detected via phase locking, and if phase locking did not improve with time in noisy environments, FM detection in noise, like STM detection, should not improve when the FM signal is delayed from the noise onset. This would also be consistent with the fact that adaptation to noise in natural word recognition [which partly relies on temporal fine structure (TFS) encoding, Hopkins and Moore (2009)] is less than adaptation to noise in vocoded word recognition.

Previous studies have shown that FM detection can improve when the FM signal is delayed from the carrier onset (Bryne et al., 2013) or when preceded by a precursor (Mesik and Wojtczak, 2020). However, the detection of FM in Byrne et al. (2013) and Mesik and Wojtczak (2020) could be due to the detection of AM cues via a mechanism known as FM-to-AM conversion (Zwicker, 1970; Moore and Sek, 1992, 1994,1995; Whiteford and Oxenham, 2015; Moore et al., 2019; Mesik and Wojtczak, 2020; Whiteford et al., 2020) since these are high-frequency carriers frequency modulated at fast rates. For this reason, the improvement in FM detection reported by Byrne et al. (2013) and Mesik and Wojtczak (2020) is unlikely due to enhanced encoding using phase locking. It would be interesting to experimentally test whether there is noise adaptation in FM detection when FM cannot be detected based on AM cues (e.g., when the frequency of the carrier signal and the modulation rate are low, or when an FM signal is also amplitude modulated to prevent detecting FM by using AM cues). If STM sensitivity is related to FM detection under such conditions, as suggested by Bernstein et al. (2013), we would expect an absence of adaptation to noise in FM detection, suggesting that both, STM sensitivity and FM detection, rely on the encoding of TFS.

### *5.2.6. Lack of correlation between adaptation to noise in word recognition and modulation detection*

The observed lack of correlation between adaptation to noise in word recognition and modulation detection (**Chapter 4**) could be due to multiple causes that could be investigated in futures studies. First, it could reflect the limited variability in our sample, which only included normal-hearing (NH) participants. Previous studies that reported significant correlations between STM sensitivity and speech perception in noise (Bernstein et al., 2013, 2016; Zaar et al., 2023) included hearing-impaired (HI) listeners in their sample. The inclusion of HI participants could potentially yield higher correlations, as the sample would have higher variability in modulation detection thresholds, speech reception thresholds (SRTs), and adaptation.

Second, it is possible that noise adaptation in word recognition relates to an enhancement of speech cues different from sustained spectral or temporal envelopes. The other speech cues in question could be transients, such as formant transitions, which have been reported to be essential for phoneme discrimination (Liberman et al., 1954; Varnet et al., 2013). Behavioral reverse correlation techniques (Varnet et al. 2013, 2015; Varnet and Lorenzi, 2022; Osses and Varnet, 2024) could help elucidating whether the cues used for speech recognition differ as the speech is delayed from the noise onset.

Third, listeners may use different strategies to perform speech recognition and modulation detection tasks. In a three-interval, three-alternative, forced-choice modulation detection task, like the one used here, participants are asked to discriminate between a modulated stimulus (the target) and two unmodulated (or standard) stimuli that are presented sequentially. In contrast, the speech recognition process comprises the discrimination and selection among competing lexical items, which are activated by the acoustic presentation of a word (Luce and Pisoni, 1998). If the presentation of a word activates lexical items with similar spectrotemporal modulations, participants must discriminate and select the correct word from a set of similar word candidates. Based on this, speech recognition may be more strongly related to modulation discrimination than to modulation detection since both modulation discrimination and speech recognition involve distinguishing between two or more modulation patterns, rather than detecting if a stimulus is modulated. If that were the case, then adaptation to noise in speech recognition might be more closely related to adaptation to noise in modulation discrimination than modulation detection. However, it is worth noting that Davies-Venn et al. (2015) found speech-in-noise recognition to be similarly correlated to modulation detection and modulation discrimination. On the other hand, since word recognition is also affected by lexical processes such as the number of lexically analogous words and the frequency of these similar lexical items (Luce and Pisoni, 1998), any improvement in the encoding of temporal and spectral envelopes, which is presumably the key factor in the improvement of modulation detection, may be blurred by the influence of these lexical aspects resulting in non-significant correlations between noise adaptation values in the two tasks. This suggests that a correlation between adaptation to noise in speech recognition and SM or TM detection (or discrimination) may emerge when speech recognition does not depend on lexical factors, e.g., when the speech recognition involves nonsense syllables or pseudo-words.

Fourth, as discussed in **Chapter 4**, there are other phenomena besides neural DRA that may contribute to adaptation to noise. The lack of correlation between noise adaptation in word recognition and modulation detection may be due to these phenomena. For example, adaptation to noise in word recognition could be due an attentional enhancement, whereas the adaptation to noise in modulation detection could be due to some sort of modulation masking (Gallun and Hafter, 2006). It may be of interest for future studies to investigate the effect of these confounding factors.

# CONCLUSIONS

This thesis was aimed at investigating if human auditory frequency selectivity and the sensitivity to spectral, temporal, and spectrotemporal modulations are affected by noise adaptation processes. Another aim was to investigate whether noise adaptation in speech recognition relates to improvements in the sensitivity to temporal, spectral, and/or spectrotemporal speech cues.

The main conclusions are:

(1) In noise backgrounds, human auditory frequency selectivity and the sensitivity to spectral and temporal modulations can change over time.

(2) Masker duration does not affect behavioral estimates of human cochlear tuning at 4 kHz. This suggests that adaptation processes do not affect tuning estimates at high frequencies. In contrast, short maskers produce broader tuning estimates than long maskers at 500 Hz. This indicates that tuning estimates at low frequencies may be affected by adaptation or other factors and should be interpreted with caution.

(3) Contralateral noise precursors hardly affect human auditory frequency selectivity at low (500 Hz) or high frequencies (4 kHz). By contrast, ipsilateral and bilateral precursors increase frequency selectivity at low frequencies and decrease it at high frequencies. While seemingly different, the two patterns are broadly consistent with efferent inhibition of cochlear gain by a precursor-activated MOCR.

(4) The recognition of words in noise improves as words are delayed in the noise. The sensitivity to spectral and temporal modulations, but not spectrotemporal modulations, improves when the modulated signals are delayed in the noise. This suggests that adaptation to noise in word recognition may be related to improvements in the neural encoding of spectral or temporal speech envelopes, but not to an improved sensitivity to its spectrotemporal modulations.

# REFERENCES


Aguilar, E., Eustaquio-Martín, A., and Lopez-Poveda, E. A. (2013). Contralateral Efferent Reflex Effects on Threshold and Suprathreshold Psychoacoustical Tuning Curves at Low and High Frequencies. *Journal of the Association for Research in Otolaryngology*, *14*(3), 341–357. https://doi.org/10.1007/s10162-013-0373-4

Ainsworth, W. A., and Meyer, G. F. (1994). Recognition of plosive syllables in noise: Comparison of an auditory model with human performance. *The Journal of the Acoustical Society of America*, *96*(2), 687–694. https://doi.org/10.1121/1.410306

Ainsworth, W. A., and Millar, J. B. (1972). The Effect of Relative Formant Amplitude on the Perceived Identity of Synthetic Vowels. *Language and Speech*, *15*(4), 328–341. https://doi.org/10.1177/002383097201500403

Almishaal, A., Bidelman, G. M., and Jennings, S. G. (2017). Notched-noise precursors improve detection of low-frequency amplitude modulation. *The Journal of the Acoustical Society of America*, *141*(1), 324–333. https://doi.org/10.1121/1.4973912

American National Standards Institute (1996). S3.6 Specification for audiometers. New York: American National Standards Institute.

Archer-Boyd, A. W., Southwell, R. V., Deeks, J. M., Turner, R. E., and Carlyon, R. P. (2018). Development and validation of a spectrotemporal processing test for cochlear-implant listeners. *The Journal of the Acoustical Society of America*, 144(5), 2983–2997. https://doi.org/10.1121/1.5079636

Aronoff, J. M., and Landsberger, D. M. (2013). The development of a modified spectral ripple test. *The Journal of the Acoustical Society of America*, *134*(2), 217–222. https://doi.org/10.1121/1.4813802

Azadpour, M., and McKay, C. M. (2012). A Psychophysical Method for Measuring Spatial Resolution in Cochlear Implants. *Journal of the Association for Research in Otolaryngology*, *13*(1), 145–157. https://doi.org/10.1007/s10162-011-0294-z

Backus, B. C., and Guinan, J. J. (2006). Time-course of the human medial olivocochlear reflex. *The Journal of the Acoustical Society of America*, *119*(5), 2889–2904. https://doi.org/10.1121/1.2169918

Bacon, S. P., and Savel, S. (2004). Temporal effects in simultaneous masking with on- and off-frequency noise maskers: Effects of signal frequency and masker level. *The Journal of the Acoustical Society of America*, *115*(4), 1674–1683. https://doi.org/10.1121/1.1689344

Bacon, S.P., and Takahashi, G. A. (1992). Overshoot in normal-hearing and hearing-impaired subjects. *The Journal of the Acoustical Society of America, 91*(5), 2865–2871. https://doi.org/10.1121/1.402967

Bacon, S. P., and Jesteadt, W. (1987). Effects of pure-tone forward masker duration on

psychophysical measures of frequency selectivity. *The Journal of the Acoustical Society of America*, *82*(6), 1925–1932. https://doi.org/10.1121/1.395688

Békésy G. von. (1960). *Experiments in hearing*. McGraw-Hill. New York.

Ben-David, B. M., Avivi-Reich, M., and Schneider, B. A. (2016). Does the degree of linguistic experience (native versus nonnative) modulate the degree to which listeners can benefit from a delay between the onset of the maskers and the onset of the target speech? *Hearing Research*, *341*, 9–18. https://doi.org/10.1016/j.heares.2016.07.016

Ben-David, B. M., Tse, V. Y. Y., and Schneider, B. A. (2012). Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? *Hearing Research*, *290*(1–2), 55–63. https://doi.org/10.1016/j.heares.2012.04.022

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*: *Series B (Methodological)*. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berlin, C. I., Hood, L. J., Hurley, A. E., Wen, H., and Kemp, D. T. (1995). Binaural noise suppresses linear click-evoked otoacoustic emissions more than ipsilateral or contralateral noise. *Hearing Research*, *87*(1–2), 96–103. https://doi.org/10.1016/0378-5955(95)00082-F

Bernstein, J. G. W., Danielsson, H., Hällgren, M., Stenfelt, S., Rönnberg, J., and Lunner, T. (2016). Spectrotemporal Modulation Sensitivity as a Predictor of Speech-Reception Performance in Noise with Hearing Aids. *Trends in Hearing*, *20*, 1–17. https://doi.org/10.1177/2331216516670387

Bernstein, J. G. W., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., and Leek, M. R. (2013). Spectrotemporal Modulation Sensitivity as a Predictor of Speech Intelligibility for Hearing-Impaired Listeners. *Journal of the American Academy of Audiology*, *24*(04), 293–306. https://doi.org/10.3766/jaaa.24.4.5

Blackburn, C. C., and Sachs, M. B. (1990). The representations of the steady-state vowel sound /ɛ/ in the discharge patterns of cat anteroventral cochlear nucleus neurons. *Journal of Neurophysiology*, *63*(5), 1191–1212. https://doi.org/10.1152/jn.1990.63.5.1191

Borg, E. (1982). Time Course of the Human Acoustic Stapedius Reflex: A Comparison of Eight Different Measures in Normal-hearing Subjects. *Scandinavian Audiology*, *11*(4), 237–242. https://doi.org/10.3109/01050398209087473

Borg, E., Counter, A., Engström, B., Linde, G., and Marklund, K. (1990). Stapedius reflex thresholds in relation to tails of auditory nerve fiber frequency tuning curves. Brain Research, *506*(1), 79–84. https://doi.org/10.1016/0006-8993(90)91201-Q

Brown, M. C. (1989). Morphology and response properties of single olivocochlear fibers in the guinea pig. *Hearing Research*, *40*(1–2), 93–109. https://doi.org/10.1016/0378-5955(89)90103-2

Byrne, A. J., Viemeister, N. F., and Stellmack, M. A. (2013). The effects of unmodulated carrier fringes on the detection of frequency modulation. *The Journal of the Acoustical Society of America*, *133*(2), 998–1003. https://doi.org/10.1121/1.4773353

Cárdenas, M.R. and Marrero, V. (1994) Cuaderno de logoaudiometría. Madrid: Universidad Nacional de Educación a Distancia.

Carlyon, R. P., and White, L. J. (1992). Effect of signal frequency and masker level on the frequency regions responsible for the overshoot effect. *The Journal of the Acoustical Society of America*, *91*(2), 1034–1041. https://doi.org/10.1121/1.402629

Cervera, T., and Ainsworth, W. A. (2005). Effects of preceding noise on the perception of voiced plosives. *Acta Acustica united with Acustica*, *91* (1), 132–144.

Cervera, T., and Gonzalez-Alvarez, J. (2007). Temporal effects of preceding band-pass and band-stop noise on the recognition of voiced stops. *Acta Acustica United with Acustica*, *93*(6), 1036–1045.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (1999). Spectrotemporal modulation transfer functions and speech intelligibility. *The Journal of the Acoustical Society of America*, *106*(5), 2719–2732. https://doi.org/10.1121/1.428100

Church, G. T., and Cudahy, E. A. (1984). The Time Course of the Acoustic Reflex. *Ear and Hearing*, *5*(4), 235–242. https://doi.org/10.1097/00003446-198407000-00008

Cooper, N. P. (1996). Two-tone suppression in cochlear mechanics. *The Journal of the Acoustical Society of America*, *99*(5), 3087–3098. https://doi.org/10.1121/1.414795

Cooper, N. P., and Guinan, J. J. (2003). Separate mechanical processes underlie fast and slow effects of medial olivocochlear efferent activity. *The Journal of Physiology*, *548*(1), 307–312. https://doi.org/10.1113/jphysiol.2003.039081

Cooper, N. P., and Guinan, J. J. (2006). Efferent-mediated control of basilar membrane motion. *Journal of Physiology*, *576*(1), 49–54. https://doi.org/10.1113/jphysiol.2006.114991

Cooper, N. P., and Rhode, W. S. (1995). Nonlinear mechanics at the apex of the guinea-pig cochlea. *Hearing Research*, *82*(2), 225–243. https://doi.org/10.1016/0378-5955(94)00180-X

Costalupes, J. A., Young, E. D., and Gibson, D. J. (1984). Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat. *Journal of Neurophysiology*, *51*(6), 1326–1344. https://doi.org/10.1152/jn.1984.51.6.1326

Dau, T., Püschel, D., and Kohlrausch, A. (1996). A quantitative model of the ‘“effective”’ signal processing in the auditory system. II. Simulations and measurements. *The Journal of the Acoustical Society of America*, *99*(6), 3623–3631. https://doi.org/10.1121/1.414960

Davies-Venn, E., Nelson, P., and Souza, P. (2015). Comparing auditory filter bandwidths, spectral ripple modulation detection, spectral ripple discrimination, and speech recognition: Normal and impaired hearing. *The Journal of the Acoustical Society of America*, *138*(1), 492–503. https://doi.org/10.1121/1.4922700

Davis, H. (1983). An active process in cochlear mechanics. *Hearing Research*, *9*(1), 79–90. https://doi.org/10.1016/0378-5955(83)90136-3

de Boer, E. (1962). Note on the Critical Bandwidth. *The Journal of the Acoustical Society of*

*America, 34*(7): 985–986. https://doi.org/10.1121/1.1918235

de Boer, E. (1966). Intensity Discrimination of Fluctuating Signals. *The Journal of the Acoustical Society of America*, *40*(3): 552–560. https://doi.org/10.1121/1.1910118

Dean, I., Harper, N. S., and McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nature Neuroscience*, *8*(12), 1684–1689. https://doi.org/10.1038/nn1541

Dean, I., Robinson, B. L., Harper, N. S., and McAlpine, D. (2008). Rapid neural adaptation to sound level statistics. *Journal of Neuroscience*, *28*(25), 6430–6438. https://doi.org/10.1523/JNEUROSCI.0470-08.2008

DeRoy Milvae, K., and Strickland, E. A. (2018). Psychoacoustic measurements of ipsilateral cochlear gain reduction as a function of signal frequency. *The Journal of the Acoustical Society of America*, *143*(5), 3114–3125. https://doi.org/10.1121/1.5038254

DeRoy Milvae, K., and Strickland, E. A. (2021). Behavioral Measures of Cochlear Gain Reduction Depend on Precursor Frequency, Bandwidth, and Level. *Frontiers in Neuroscience*, *15*(October), 1–14. https://doi.org/10.3389/fnins.2021.716689

Dolan, D. F., Guo, M. H., and Nuttall, A. L. (1997). Frequency-dependent enhancement of basilar membrane velocity during olivocochlear bundle stimulation. *The Journal of the Acoustical Society of America*, *102*(6), 3587–3596. https://doi.org/10.1121/1.421008

Drullman, R., Festen, J. M., and Houtgast, T. (1996). Effect of temporal modulation reduction on spectral contrasts in speech. *The Journal of the Acoustical Society of America*, *99*(4), 2358–2364. https://doi.org/10.1121/1.415423

Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, *95*(5), 2670–2680. https://doi.org/10.1121/1.409836

Eddins, D. A., and Bero, E. M. (2007). Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region. *The Journal of the Acoustical Society of America*, *121*(1), 363–372. https://doi.org/10.1121/1.2382347

Elliot, L. L. (1965). Changes in the simultaneous masked threshold of brief tones. *The Journal of the Acoustical Society of America, 38*(5), 738-746. https://doi.org/10.1121/1.1909798

Elliott, T. M., and Theunissen, F. E. (2009). The Modulation Transfer Function for Speech Intelligibility. *PLoS Computational Biology*, *5*(3), e1000302. https://doi.org/10.1371/journal.pcbi.1000302

Escabí, M. A., and Schreiner, C. E. (2002). Nonlinear Spectrotemporal Sound Analysis by Neurons in the Auditory Midbrain. *Journal of Neuroscience*, *22*(10), 4114–4131. https://doi.org/10.1523/jneurosci.22-10-04114.2002

Eustaquio-Martín, A., and Lopez-Poveda, E. A. (2011). Isoresponse versus isoinput estimates of cochlear filter tuning. *Journal of the Association for Research in Otolaryngology*, *12*(3), 281–299. https://doi.org/10.1007/s10162-010-0252-1

Evans, E.F. and Wilson, J.P. (1973). Frequency selectivity of the cochlea, in *Basic Mechanisms*

*in Hearing,* edited by A.R. Moller, Academic Press, NY, 519–551.

Evans, E. F. (1975). Auditory frequency selectivity and the cochlear nerve, in *Facts and Models of Hearing*, edited by E. Zwicker and E. Terhard, Springer-Verlag, Berlin, 118–132.

Evans, E. F. (2001). Latest comparisons between physiological and behavioral frequency selectivity. In *Physiological and psychophysical bases of auditory function*, edited by J. Breebaart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs, R. Schoonhoven, Shaker, Maastricht, 382–387.

Farhadi, A., Jennings, S. G., Strickland, E. A., and Carney, L. H. (2023). Subcortical auditory model including efferent dynamic gain control with inputs from cochlear nucleus and inferior colliculus. *The Journal of the Acoustical Society of America*, *154*(6), 3644–3659. https://doi.org/10.1121/10.0022578

Feeney, M. P., Keefe, D. H., Hunter, L. L., Fitzpatrick, D. F., Garinis, A. C., Putterman, D. B., and McMillan, G. P. (2017). Normative Wideband Reflectance, Equivalent Admittance at the Tympanic Membrane, and Acoustic Stapedius Reflex Threshold in Adults. *Ear and Hearing*, *38*(3), e142–e160. https://doi.org/10.1097/AUD.0000000000000399

Ferry, R. T., and Meddis, R. (2007). A computer model of medial efferent suppression in the mammalian auditory system. *The Journal of the Acoustical Society of America*, *122*(6), 3519–3526. https://doi.org/10.1121/1.2799914

Fletcher, H. (1940). Auditory Patterns. *Reviews of Modern Physics*, *12*(1), 47–65. https://doi.org/10.1103/RevModPhys.12.47

Fletcher, M. D., Boer, J. De, and Krumbholz, K. (2013). Is Overshoot Caused by an Efferent Reduction in Cochlear Gain? In *Basic Aspects of Hearing* editedy by B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, and H. E. Gockel, Springer, NY, 65–72. https://doi.org/10.1007/978-1-4614-1590-9

Francis, N. A., and Guinan, J. J. (2010). Acoustic stimulation of human medial olivocochlear efferents reduces stimulus-frequency and click-evoked otoacoustic emission delays: Implications for cochlear filter bandwidths. *Hearing Research*, *267*(1–2), 36–45. https://doi.org/10.1016/j.heares.2010.04.009

Gallun, F. J., and Hafter, E. R. (2006). Amplitude modulation sensitivity as a mechanism for increment detection. *The Journal of the Acoustical Society of America*, *119*(6), 3919–3930. https://doi.org/10.1121/1.2200136

Gibson, D. J., Young, E. D., and Costalupes, J. A. (1985). Similarity of dynamic range adjustment in auditory nerve and cochlear nuclei. *Journal of Neurophysiology*, *53*(4), 940–958. https://doi.org/10.1152/jn.1985.53.4.940

Glasberg, B. R., and Moore, B. C. J. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *The Journal of the Acoustical Society of America*, *79*(4), 1020–1033. https://doi.org/10.1121/1.393374

Glasberg, B. R., and Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1–2), 103–138. https://doi.org/10.1016/0378-

5955(90)90170-T

Goode, R. L., Killion, M., Nakamura, K., and Nishihara, S. (1994). New knowledge about the function of the human middle ear: development of an improved analog model. *The American Journal of Otology*, *15*(2), 145–154. http://www.ncbi.nlm.nih.gov/pubmed/8172293

Grange, J., Zhang, M., and Culling, J. (2022). The Role of Efferent Reflexes in the Efficient Encoding of Speech by the Auditory Nerve. *The Journal of Neuroscience*, *42*(36), 6907–6916. https://doi.org/10.1523/JNEUROSCI.2220-21.2022

Guinan, J. J. (2006). Olivocochlear Efferents: Anatomy, Physiology, Function, and the Measurement of Efferent Effects in Humans. *Ear and Hearing*, *27*(6), 589–607. https://doi.org/10.1097/01.aud.0000240507.83072.e7

Guinan, J. J. (2010). Cochlear efferent innervation and function. *Current Opinion in Otolaryngology and Head and Neck Surgery*, *18*(5), 447–453. https://doi.org/10.1097/MOO.0b013e32833e05d6

Henry, B. A., Turner, C. W., and Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. *The Journal of the Acoustical Society of America*, *118*(2), 1111–1121. https://doi.org/10.1121/1.1944567

Hopkins, K., and Moore, B. C. J. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *The Journal of the Acoustical Society of America*, *125*(1), 442–446. https://doi.org/10.1121/1.3037233

Hothorn, T., Bretz, F., Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3),346-63. https://doi.org/10.1002/bimj.200810425

Hung, I. J., and Dallos, P. (1972). Study of the Acoustic Reflex in Human Beings. I. Dynamic Characteristics. *The Journal of the Acoustical Society of America*, *52*(4B), 1168–1180. https://doi.org/10.1121/1.1913229

Isarangura, S., Eddins, A. C., Ozmeral, E. J., and Eddins, D. A. (2019). The Effects of Duration and Level on Spectral Modulation Perception. *Journal of Speech, Language, and Hearing Research*, *62*(10), 3876–3886. https://doi.org/10.1044/2019_JSLHR-H-18-0449

Jennings, S. G. (2021). The role of the medial olivocochlear reflex in psychophysical masking and intensity resolution in humans: a review. *Journal of Neurophysiology*, *125*(6), 2279–2308. https://doi.org/10.1152/jn.00672.2020

Jennings, S. G., and Strickland, E. A. (2012a). Auditory filter tuning inferred with short sinusoidal and notched-noise maskers. *The Journal of the Acoustical Society of America*, *132*(4), 2497–2513. https://doi.org/10.1121/1.4746029

Jennings, S. G., and Strickland, E. A. (2012b). Evaluating the effects of olivocochlear feedback on psychophysical measures of frequency selectivity. *The Journal of the Acoustical Society of America*, *132*(4), 2483–2496. https://doi.org/10.1121/1.4742723

Jennings, S. G., Heinz, M. G., and Strickland, E. A., (2011). Evaluating adaptation and olivocochlear efferent feedback as potential explanations of psychophysical overshoot. *Journal of the Association for Research in Otolaryngology*, *12*(3), 345-360. https://doi.org/10.1007/s10162-011-0256-5

Jennings, S. G., Strickland, E. A., and Heinz, M. G. (2009). Precursor effects on behavioral estimates of frequency selectivity and gain in forward masking. *The Journal of the Acoustical Society of America*, *125*(4), 2172–2181. https://doi.org/10.1121/1.3081383

Kassambara, A. (2021). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.0. https://CRAN.R-project.org/package=rstatix

Kawase, T., and Liberman, M. C. (1993). Antimasking effects of the olivocochlear reflex. I. Enhancement of compound action potentials to masked tones. *Journal of Neurophysiology*, *70*(6), 2519–2532. https://doi.org/10.1152/jn.1993.70.6.2519

Kawase, T., Ogura, M., Hidaka, H., Sasaki, N., Suzuki, Y., and Takasaka, T. (2000). Effects of contralateral noise on measurement of the psychophysical tuning curve. *Hearing Research*, *142*(1–2), 63–70. https://doi.org/10.1016/S0378-5955(00)00010-1

Keefe, D. H., Fitzpatrick, D., Liu, Y.-W., Sanford, C. A., and Gorga, M. P. (2010). Wideband acoustic-reflex test in a test battery to predict middle-ear dysfunction. *Hearing Research*, *263*(1), 52–65. https://doi.org/10.1016/j.heares.2009.09.008

Kidd, G., Mason, C. R., and Feth, L. L. (1984). Temporal integration of forward masking in listeners having sensorineural hearing loss. *The Journal of the Acoustical Society of America*, *75*(3), 937–944. https://doi.org/10.1121/1.390558

Kluk, K., and Moore, B. C. J. (2004). Factors affecting psychophysical tuning curves for normally hearing subjects. *Hearing Research*, *194*(1–2), 118–134. https://doi.org/10.1016/j.heares.2004.04.012

Krull, V., and Strickland, E. A. (2008). The effect of a precursor on growth of forward masking. *The Journal of the Acoustical Society of America*, *123*(6), 4352–4357. https://doi.org/10.1121/1.2912440

Leek, M. R., Dorman, M. F., and Summerfield, Q. (1987). Minimum spectral contrast for vowel identification by normalhearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *81*(1), 148–154. https://doi.org/10.1121/1.395024

Leek, M. R., and Summers, V. (1996). Reduced frequency selectivity and the preservation of spectral contrast in noise. *The Journal of the Acoustical Society of America*, *100*(3), 1796–1806. https://doi.org/10.1121/1.415999

Leschke, J., Rodriguez Orellana, G., Shera, C. A., and Oxenham, A. J. (2022). Auditory filter shapes derived from forward and simultaneous masking at low frequencies: Implications for human cochlear tuning. *Hearing Research*, *420*, 108500. https://doi.org/10.1016/j.heares.2022.108500

Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477. https://doi.org/10.1121/1.1912375

Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied*, *68*(8), 1–13. https://doi.org/10.1037/h0093673

Liberman, M. C., and Brown, M. C. (1986). Physiology and anatomy of single olivocochlear neurons in the cat. *Hearing Research*, *24*(1), 17–36. https://doi.org/10.1016/0378-5955(86)90003-1

Liberman, M. C., and Guinan, J. J. (1998). Feedback control of the auditory periphery. *Journal of Communication Disorders*, *31*(6), 471–483. https://doi.org/10.1016/S0021-9924(98)00019-7

Liberman, L. D., and Liberman, M. C. (2019). Cochlear Efferent Innervation Is Sparse in Humans and Decreases with Age. *The Journal of Neuroscience*, *39*(48), 9560–9569. https://doi.org/10.1523/JNEUROSCI.3004-18.2019

Lilaonitkul, W., and Guinan, J. J. (2009a). Human Medial Olivocochlear Reflex: Effects as Functions of Contralateral, Ipsilateral, and Bilateral Elicitor Bandwidths. *Journal of the Association for Research in Otolaryngology*, *10*(3), 459–470. https://doi.org/10.1007/s10162-009-0163-1

Lilaonitkul, W., and Guinan, J. J. (2009b). Reflex Control of the Human Inner Ear: A Half-Octave Offset in Medial Efferent Feedback That Is Consistent With an Efferent Role in the Control of Masking. *Journal of Neurophysiology*, *101*(3), 1394–1406. https://doi.org/10.1152/jn.90925.2008

Lilaonitkul, W., and Guinan, J. J. (2012). Frequency tuning of medial-olivocochlear-efferent acoustic reflexes in humans as functions of probe frequency. *Journal of Neurophysiology*, *107*(6), 1598–1611. https://doi.org/10.1152/jn.00549.2011

Liu, C., and Eddins, D. A. (2008). Effects of spectral modulation filtering on vowel identification. *The Journal of the Acoustical Society of America*, *124*(3), 1704–1715. https://doi.org/10.1121/1.2956468

Lopez-Poveda, E. A. (2018). Olivocochlear Efferents in Animals and Humans: From Anatomy to Clinical Relevance. *Frontiers in Neurology*, 9(MAR), 1–18. https://doi.org/10.3389/fneur.2018.00197

Lopez-Poveda, E. A. (2005). Spectral Processing by the Peripheral Auditory System: Facts and Models. In *International Review of Neurobiology*, *70*(5), 7–48. https://doi.org/10.1016/S0074-7742(05)70001-5

Lopez-Poveda, E. A., and Eustaquio-Martín, A. (2013). On the controversy about the sharpness of human cochlear tuning. *Journal of the Association for Research in Otolaryngology*, *14*(5), 673–686. https://doi.org/10.1007/s10162-013-0397-9

Lopez-Poveda, E. A., Eustaquio-Martín, A., Stohl, J. S., Wolford, R. D., Schatzer, R., and Wilson, B. S. (2016). A binaural cochlear implant sound coding strategy inspired by the contralateral medial olivocochlear reflex. *Ear and Hearing*, *37*(3), e138–e148. https://doi.org/10.1097/AUD.0000000000000273

Lopez-Poveda, E. A., and Meddis, R. (2001). A human nonlinear cochlear filterbank. *The Journal of the Acoustical Society of America*, *110*(6), 3107–3118. https://doi.org/10.1121/1.14161

Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (2003). Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing. *The Journal of the Acoustical Society of America*, *113*(2), 951–960. https://doi.org/10.1121/1.1534838

Luce, P. A., and Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Magits, S., Moncada-Torres, A., Van Deun, L., Wouters, J., van Wieringen, A., and Francart, T. (2019). The effect of presentation level on spectrotemporal modulation detection. *Hearing Research*, *371*, 11–18. https://doi.org/10.1016/j.heares.2018.10.017

Marrufo-Pérez, M. I., Sturla-Carreto, D. del P., Eustaquio-Martín, A., and Lopez-Poveda, E. A. (2020). Adaptation to Noise in Human Speech Recognition Depends on Noise-Level Statistics and Fast Dynamic-Range Compression. *The Journal of Neuroscience*, *40*(34), 6613–6623. https://doi.org/10.1523/JNEUROSCI.0469-20.2020

Marrufo-Pérez, M. I., Johannesen, P. T., and Lopez-Poveda, E. A. (2021). Correlation and Reliability of Behavioral and Otoacoustic-Emission Estimates of Contralateral Medial Olivocochlear Reflex Strength in Humans. *Frontiers in Neuroscience*, *15*, 1–18. https://doi.org/10.3389/fnins.2021.640127

Marrufo-Pérez, M. I., Eustaquio-Martín, A., Fumero, M. J., Gorospe, J. M., Polo, R., Gutiérrez Revilla, A., and Lopez-Poveda, E. A. (2019). Adaptation to noise in amplitude modulation detection without the medial olivocochlear reflex. *Hearing Research*, *377*, 133–141. https://doi.org/10.1016/j.heares.2019.03.017

Marrufo-Pérez, M. I., Eustaquio-Martín, A., López-Bascuas, L. E., and Lopez-Poveda, E. A. (2018b). Temporal Effects on Monaural Amplitude-Modulation Sensitivity in Ipsilateral, Contralateral and Bilateral Noise. *Journal of the Association for Research in Otolaryngology*, *19*(2), 147–161. https://doi.org/10.1007/s10162-018-0656-x

Marrufo-Pérez, M. I., Eustaquio-Martín, A., and Lopez-Poveda, E. A. (2018a). Adaptation to Noise in Human Speech Recognition Unrelated to the Medial Olivocochlear Reflex. *The Journal of Neuroscience*, *38*(17), 4138–4145. https://doi.org/10.1523/JNEUROSCI.0024-18.2018

Marrufo-Pérez, M. I., and Lopez-Poveda, E. A. (2022). Adaptation to noise in normal and impaired hearing. *The Journal of the Acoustical Society of America*, *151*(3), 1741–1753. https://doi.org/10.1121/10.0009802

May, B. J., and Sachs, M. B. (1992). Dynamic range of neural rate responses in the ventral cochlear nucleus of awake cats. *Journal of Neurophysiology*, *68*(5), 1589–1602. https://doi.org/10.1152/jn.1992.68.5.1589

McGregor, K. D., Flamme, G. A., Tasko, S. M., Deiters, K. K., Ahroon, W. A., Themann, C. L., and

Murphy, W. J. (2018). Acoustic reflexes are common but not pervasive: evidence using a diagnostic middle ear analyser. *International Journal of Audiology*, *57*, S42–S50. https://doi.org/10.1080/14992027.2017.1416189

Mehraei, G., Gallun, F. J., Leek, M. R., and Bernstein, J. G. W. (2014). Spectrotemporal modulation sensitivity for hearing-impaired listeners: Dependence on carrier center frequency and the relationship to speech intelligibility. *The Journal of the Acoustical Society of America*, *136*(1), 301–316. https://doi.org/10.1121/1.4881918

Mesik, J., and Wojtczak, M. (2020). Effects of noise precursors on the detection of amplitude and frequency modulation for tones in noise. *The Journal of the Acoustical Society of America*, *148*(6), 3581–3597. https://doi.org/10.1121/10.0002879

Moore, B. C. J., Mariathasan, S., and Sek, A. P. (2019). Effects of Age and Hearing Loss on the Discrimination of Amplitude and Frequency Modulation for 2- and 10-Hz Rates. *Trends in Hearing*, *23*, 1–12. https://doi.org/10.1177/2331216519853963

Moore, B. C. J., and Sek, A. (1992). Detection of combined frequency and amplitude modulation. *The Journal of the Acoustical Society of America*, *92*(6), 3119–3131. https://doi.org/10.1121/1.404208

Moore, B. C. J., and Sek, A. (1994). Effects of carrier frequency and background noise on the detection of mixed modulation. *The Journal of the Acoustical Society of America*, *96*(2), 741–751. https://doi.org/10.1121/1.410312

Moore, B. C. J., and Sek, A. (1995). Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation). *The Journal of the Acoustical Society of America*, *97*(4), 2468–2478. https://doi.org/10.1121/1.411967

Moore, B. C. J., and Sek, A. (1996).Detection of frequency modulation at low modulatino rates: evidence for a mechanism based on phase locking. *The Journal of the Acoustical Society of America*, *100*(4), 2320–2331. https://doi.org/10.1121/1.417941

Murugasu, E., and Russell, I. J. (1996). The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea. *The Journal of Neuroscience*, *16*(1), 325–332. https://doi.org/10.1523/JNEUROSCI.16-01-00325.1996

Narayan, S.S., Temchin, A.N., Recio, A. and Ruggero, M.A. (1998) Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae. *Science, 282*(5395), 1882–1884. https://doi.org/10.1126/science.282.5395.1882

Narne, V. K., Prabhu, P., Van Dun, B., and Moore, B. C. J. (2018). Ripple glide direction discrimination and its relationship to frequency selectivity estimated using notched noise. *Acta Acustica United with Acustica*, *104*(6), 1063–1074. https://doi.org/10.3813/AAA.919272

Narne, V. K., Sharma, M., Van Dun, B., Bansal, S., Prabhu, L., and Moore, B. C. J. (2016). Effects of spectral smearing on performance of the spectral ripple and spectrotemporal ripple tests. *The Journal of the Acoustical Society of America*, *140*(6), 4298–4306. https://doi.org/10.1121/1.4971419

Neff, D. L. (1985). Stimulus parameters governing confusion effects in forward masking. *The*

*Journal of the Acoustical Society of America*, *78*(6), 1966–1976. https://doi.org/10.1121/1.392653

Neff, D. L. (1986). Confusion effects with sinusoidal and narrow-band noise forward maskers. *The Journal of the Acoustical Society of America*, *79*(5), 1519–1529. https://doi.org/10.1121/1.393678

Nelson, D. A., and Freyman, R. L. (1984). Broadened forward-masked tuning curves from intense masking tones: Delay-time and probe-level manipulations. *The Journal of the Acoustical Society of America*, *75*(5), 1570–1577. https://doi.org/10.1121/1.390866

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (2001). A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *110*(4), 2045–2064. https://doi.org/10.1121/1.1404439

Oertel, D., Bal, R., Gardner, S. M., Smith, P. H., and Joris, P. X. (2000). Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22), 11773–11779. https://doi.org/10.1073/pnas.97.22.11773

Oetjen, A., and Verhey, J. L. (2015). Spectrotemporal modulation masking patterns reveal frequency selectivity. *The Journal of the Acoustical Society of America*, 137(2), 714–723. https://doi.org/10.1121/1.4906171

O'Loughlin, B. J., and Moore, B. C. J. (1981). Off-frequency listening: Effects on psychoacoustical tuning curves obtained in simultaneous and forward masking. *The Journal of the Acoustical Society of America*, *69*(4), 1119–1125. https://doi.org/10.1121/1.385691

Osses, A., and Varnet, L. (2024). A microscopic investigation of the effect of random envelope fluctuations on phoneme-in-noise perception. *The Journal of the Acoustical Society of America*, *155*(2), 1469–1485. https://doi.org/10.1121/10.0024469

Oxenham, A. J. (2001). Forward masking: Adaptation or integration? *The Journal of the Acoustical Society of America*, *109*(2), 732–741. https://doi.org/10.1121/1.1336501

Oxenham, A. J., and Plack, C. J. (1997). A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America*, *101*(6), 3666–3675. https://doi.org/10.1121/1.41832797

Oxenham, A. J., and Shera, C. A. (2003). Estimates of Human Cochlear Tuning at Low Levels Using Forward and Simultaneous Masking. *Journal of the Association for Research in Otolaryngology*, *4*(4), 541–554. https://doi.org/10.1007/s10162-002-3058-y

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *The Journal of the Acoustical Society of America*, *72*(6), 1788–1803. https://doi.org/10.1121/1.388652

Patterson, R. D., and Nimmo-Smith, I. (1980). Off-frequency listening and auditory-filter asymmetry. *The Journal of the Acoustical Society of America*, *67*(1), 229–245.

https://doi.org/10.1121/1.383732

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

Philibert, B., Veuillet, E., and Collet, L. (1998). Functional asymmetries of crossed and uncrossed medial olivocochlear efferent pathways in humans. *Neuroscience Letters*, *253*(2), 99–102. https://doi.org/10.1016/S0304-3940(98)00615-6

Phillips, D. P., and Hall, S. E. (1986). Spike-rate intensity functions of cat cortical neurons studied continuous noise usually caused a steepening of the slope of the tone intensity. *The Journal of the Acoustical Society of America*, *80*(1), 177–187. https://doi.org/10.1121/1.394178

Plack, C. J., Oxenham, A. J., and Drga, V. (2002). Linear and nonlinear processes in temporal masking. *Acta Acustica United with Acustica*, *88*(3), 348–358.

Ponsot, E., Varnet, L., Wallaert, N., Daoud, E., Shamma, S. A., Lorenzi, C., and Neri, P. (2021). Mechanisms of Spectrotemporal Modulation Detection for Normal- and Hearing-Impaired Listeners. *Trends in Hearing*, 25. https://doi.org/10.1177/2331216520978029

Qiu, W. W., and Stucker, F. J. (1998). Characterization of Acoustic Reflex Latency in Normal-Hearing Subjects. *Scandinavian Audiology*, *27*(1), 43–49. https://doi.org/10.1080/010503998419696

Quaranta, N., Scaringi, A., Nahum, S., and Quaranta, A. (2005). Effects of efferent acoustic reflex activation on psychoacoustical tuning curves in humans. *Acta Oto-Laryngologica*, *125*(5), 520–523. https://doi.org/10.1080/00016480510026214

Recio-Spinoso, A. and Oghalai, J.S. (2017). Mechanical tuning and amplification within the apex of the guinea pig cochlea. *Journal of Physiology, 595*, 4549–4561. https://doi.org/10.1113/JP273881

Recio-Spinoso, A., and Oghalai, J. S. (2018). Unusual mechanical processing of sounds at the apex of the Guinea pig cochlea. Hearing Research, *370*, 84–93. https://doi.org/10.1016/j.heares.2018.09.009

Rees, A., and Palmer, A. R. (1989). Neuronal responses to amplitude-modulated and pure-tone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise. *The Journal of the Acoustical Society of America*, *85*(5), 1978–1994. https://doi.org/10.1121/1.397851

Rhode, W. S. (1971). Observations of the Vibration of the Basilar Membrane in Squirrel Monkeys using the Mössbauer Technique. *The Journal of the Acoustical Society of America*, *49*(4B), 1218–1231. https://doi.org/10.1121/1.1912485

Robertson, D. (1984). Horseradish peroxidase injection of physiologically characterized afferent and efferent neurones in the guinea pig spiral ganglion. *Hearing Research*, *15*(2), 113–121. https://doi.org/10.1016/0378-5955(84)90042-X

Robles, L., and Ruggero, M. A. (2001). Mechanics of the Mammalian Cochlea. *Physiological*

*Reviews*, *81*(3), 1305–1352. https://doi.org/10.1152/physrev.2001.81.3.1305

Rodríguez, F. A., Read, H. L., and Escabí, M. A. (2010). Spectral and Temporal Modulation Tradeoff in the Inferior Colliculus. *Journal of Neurophysiology*, 103(2), 887–903. https://doi.org/10.1152/jn.00813.2009

Rosen, S., and Baker, R. J. (1994). Acoustic Reflexes in the Measurement of Auditory Filters at High Levels in Normal Listeners. *International Journal of Audiology*, *33*(1), 37–46. https://doi.org/10.3109/00206099409072953

Roverud, E., and Strickland, E. A. (2010). The time course of cochlear gain reduction measured using a more efficient psychophysical technique. *The Journal of the Acoustical Society of America*, *128*(3), 1203. https://doi.org/10.1121/1.3473695

Ruggero, M. A., and Rich, N. C. (1991). Furosemide alters organ of corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane. *Journal of Neuroscience*, *11*(4), 1057–1067. https://doi.org/10.1523/jneurosci.11-04-01057.1991

Ruggero, M. A., Robles, L., and Rich, N. C. (1992). Two-tone suppression in the basilar membrane of the cochlea: mechanical basis of auditory-nerve rate suppression. *Journal of Neurophysiology*, *68*(4), 1087–1099. https://doi.org/10.1152/jn.1992.68.4.1087

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (1997). Basilar-membrane responses to tones at the base of the chinchilla cochlea. *The Journal of the Acoustical Society of America*, *101*(4), 2151–2163. https://doi.org/10.1121/1.418265

Ruggero, M. A., and Temchin, A. N. (2005). Unexceptional sharpness of frequency tuning in the human cochlea. *Proceedings of the National Academy of Sciences*, *102*(51), 18614–18619. https://doi.org/10.1073/pnas.0509323102

Sachs, M. B., and Kiang, N. Y. S. (1968). Two-Tone Inhibition in Auditory-Nerve Fibers. *The Journal of the Acoustical Society of America*, *43*(5), 1120–1128. https://doi.org/10.1121/1.1910947

Salloom, W. B., and Strickland, E. A. (2021). The effect of broadband elicitor laterality on psychoacoustic gain reduction across signal frequency. *The Journal of the Acoustical Society of America*, *150*(4), 2817–2835. https://doi.org/10.1121/10.0006662

Salloom, W. B., Bharadwaj, H., and Strickland, E. A. (2023). The effects of broadband elicitor duration on a psychoacoustic measure of cochlear gain reduction. *The Journal of the Acoustical Society of America*, *153*(4), 2482. https://doi.org/10.1121/10.0017925

Shannon, R. V, Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with Primarily Temporal Cues. *Science*, 270(5234), 303–304. http://www.jstor.org/stable/2888543

Sheft, S., and Yost, W. A. (1990). Temporal integration in amplitude modulation detection. *The Journal of the Acoustical Society of America*, *88*(2), 796–805. https://doi.org/10.1121/1.399729

Shera, C. A., Guinan, J. J., and Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proceedings of the National*

*Academy of Sciences*, *99*(5), 3318–3323. https://doi.org/10.1073/pnas.032675099

Shera, C. A., Guinan, J. J., and Oxenham, A. J. (2010). Otoacoustic Estimation of Cochlear Tuning: Validation in the Chinchilla. *Journal of the Association for Research in Otolaryngology*, *11*(3), 343–365. https://doi.org/10.1007/s10162-010-0217-4

Small, A. M. (1959). Pure-Tone Masking. *The Journal of the Acoustical Society of America*, *32(12)*, 1619–1625. https://doi.org/10.1121/1.1907670

Strickland, E. A. (2001). The relationship between frequency selectivity and overshoot. *The Journal of the Acoustical Society of America*, *109*(5), 2062–2073. https://doi.org/10.1121/1.1357811

Summers, V., and Leek, M. R. (1994). The internal representation of spectral contrast in hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *95*(6), 3518–3528. https://doi.org/10.1121/1.409969

Sumner, C. J., Wells, T. T., Bergevin, C., Sollini, J., Kreft, H. A., Palmer, A. R., Oxenham, A. J., and Shera, C. A. (2018). Mammalian behavior and physiology converge to confirm sharper cochlear tuning in humans. *Proceedings of the National Academy of Sciences*, *115*(44), 11322–11326. https://doi.org/10.1073/pnas.1810766115

Supin, A. Y., Popov, V. V., Milekhina, O. N., and Tarakanov, M. B. (1998). Ripple density resolution for various rippled-noise patterns. *The Journal of the Acoustical Society of America*, *103*(4), 2042–2050. https://doi.org/10.1121/1.421351

Tavartkiladze, G. A., Frolenkov, G. I., and Artamasov, S. V. (1996). Ipsilateral suppression of transient evoked otoacoustic emission: Role of the medial olivocochlear system. *Acta Oto-Laryngologica*, *116*(2), 213–218. https://doi.org/10.3109/00016489609137826

van Veen, T. M., and Houtgast, T. (1985). Spectral sharpness and vowel dissimilarity. *The Journal of the Acoustical Society of America*, *77*(2), 628–634. https://doi.org/10.1121/1.391880

Varnet, L., and Lorenzi, C. (2022). Probing temporal modulation detection in white noise using intrinsic envelope fluctuations: A reverse-correlation study. *The Journal of the Acoustical Society of America*, *151*(2), 1353–1366. https://doi.org/10.1121/10.0009629

Varnet, L., Knoblauch, K., Serniclaes, W., Meunier, F., and Hoen, M. (2015). A Psychophysical Imaging Method Evidencing Auditory Cue Extraction during Speech Perception: A Group Analysis of Auditory Classification Images. *PLOS ONE*, *10*(3), e0118009. https://doi.org/10.1371/journal.pone.0118009

Varnet, L., Knoblauch, K., Meunier, F., and Hoen, M. (2013). Using auditory classification images for the identification of fine acoustic cues used in speech perception. *Frontiers in Human Neuroscience*, 7(DEC), 1–12. https://doi.org/10.3389/fnhum.2013.00865

Versteegh, C. P. C., Meenderink, S. W. F., and van der Heijden, M. (2011). Response Characteristics in the Apex of the Gerbil Cochlea Studied Through Auditory Nerve Recordings. *Journal of the Association for Research in Otolaryngology*, *12*(3), 301–316. https://doi.org/10.1007/s10162-010-0255-y

Vinay, and Moore, B. C. J. (2008). Effects of activation of the efferent system on psychophysical tuning curves as a function of signal frequency. *Hearing Research*, *240*(1–2), 93–101. https://doi.org/10.1016/j.heares.2008.03.002

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholdsa). *The Journal of the Acoustical Society of America*, *66*(5), 1364–1380. https://doi.org/10.1121/1.383531

Warr, W. B., and Guinan, J. J. (1979). Efferent innervation of the organ of corti: two separate systems. *Brain Research*, *173*(1), 152–155. https://doi.org/10.1016/0006-8993(79)91104-1

Watkins, P. V., and Barbour, D. L. (2008). Specialized neuronal adaptation for preserving input sensitivity. *Nature Neuroscience*, *11*(11), 1259–1261. https://doi.org/10.1038/nn.2201

Watkins, P. V., and Barbour, D. L. (2011). Level-tuned neurons in primary auditory cortex adapt differently to loud versus soft sounds. *Cerebral Cortex*, *21*(1), 178–190. https://doi.org/10.1093/cercor/bhq079

Wen, B., Wang, G. I., Dean, I., and Delgutte, B. (2009). Dynamic range adaptation to sound level statistics in the auditory nerve. *Journal of Neuroscience*, *29*(44), 13797–13808. https://doi.org/10.1523/JNEUROSCI.5610-08.2009

Wen, B., Wang, G. I., Dean, I., and Delgutte, B. (2012). Time course of dynamic range adaptation in the auditory nerve. *Journal of Neurophysiology*, *108*(1), 69–82. https://doi.org/10.1152/jn.00055.2012

Whiteford, K. L., and Oxenham, A. J. (2015). Using individual differences to test the role of temporal and place cues in coding frequency modulation. *The Journal of the Acoustical Society of America*, *138*(5), 3093–3104. https://doi.org/10.1121/1.4935018

Whiteford, K. L., Kreft, H. A., and Oxenham, A. J. (2020). The role of cochlear place coding in the perception of frequency modulation. *ELife*, *9*, 1–64. https://doi.org/10.7554/eLife.58468

Wicher, A. (2013). The effect of contralateral signal on distortion product otoacoustic emissions and psychophysical tuning curves at 1 and 2 kHz. *Acta Physica Polonica A*, *123*(6), 1001–1006. https://doi.org/10.12693/APhysPolA.123.1001

Wicher, A. and Moore, B. C. J. (2014). Effect of broadband and narrowband contralateral noise on psychophysical tuning curves and otoacoustic emissions. *The Journal of the Acoustical Society of America*, *135*(5), 2931–2941. https://doi.org/10.1121/1.4871358

Willmore, B. D. B., and King, A. J. (2023). Adaptation in Auditory Processing. *Physiological Reviews*, *103*(2), 1025–1058. https://doi.org/10.1152/physrev.00011.2022

Winslow, R. L., and Sachs, M. B. (1987). Effect of electrical stimulation of the crossed olivocochlear bundle on auditory nerve response to tones in noise. *Journal of Neurophysiology*, *57*(4), 1002–1021. https://doi.org/10.1152/jn.1987.57.4.1002

Winslow, R. L., and Sachs, M. B. (1988). Single-tone intensity discrimination based on auditory-

nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. *Hearing Research*, *35*(2–3), 165–189. https://doi.org/10.1016/0378-5955(88)90116-5

Wojtczak, M., Klang, A. M., and Torunsky, N. T. (2019). Exploring the Role of Medial Olivocochlear Efferents on the Detection of Amplitude Modulation for Tones Presented in Noise. *Journal of the Association for Research in Otolaryngology*, *20*(4), 395–413. https://doi.org/10.1007/s10162-019-00722-6

Xing, D., and Gong, Q. (2017). Frequency specificity and left-ear advantage of medial olivocochlear efferent modulation: A study based on stimulus frequency otoacoustic emission. *NeuroReport*, *28*(13), 775–778. https://doi.org/10.1097/WNR.0000000000000812

Yasin, I., Drga, V., and Plack, C. J. (2014). Effect of Human Auditory Efferent Feedback on Cochlear Gain and Compression. *Journal of Neuroscience*, *34*(46), 15319–15326. https://doi.org/10.1523/JNEUROSCI.1043-14.2014

Yonemoto, K., Kurauchi, N., Hamada, H., and Miura, T. (1988). Relations between frequency selectivity, temporal resolution, and speech intelligibility in sensorineural hearing-impaired subjects. *The Journal of the Acoustical Society of America*, *84*(S1), S76–S76. https://doi.org/10.1121/1.2026472

Zaar, J., Simonsen, L. B., Dau, T., and Laugesen, S. (2023). Toward a clinically viable spectrotemporal modulation test for predicting supra-threshold speech reception in hearing-impaired listeners. *Hearing Research*, *427*, 108650. https://doi.org/10.1016/j.heares.2022.108650

Zahorian, S. A., and Jagharghi, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America*, *94*(4), 1966–1982. https://doi.org/10.1121/1.407520

Zilany, M. S. A., and Carney, L. H. (2010). Power-law dynamics in an auditory-nerve model can account for neural adaptation to sound-level statistics. *Journal of Neuroscience*, *30*(31), 10380–10390. https://doi.org/10.1523/JNEUROSCI.0647-10.2010

Zwicker, E. (1965a). Temporal Effects in Simultaneous Masking and Loudness. *The Journal of the Acoustical Society of America*, *38*(1), 132–141. https://doi.org/10.1121/1.1909588

Zwicker, E. (1965b). Temporal Effects in Simultaneous Masking by White-Noise Bursts. *The Journal of the Acoustical Society of America*, *37*(4), 653–663. https://doi.org/10.1121/1.1909389

Zwicker, E. (1970). Masking and psychological excitation as consequences of the ear's frequency analysis, in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg, Sijthoff, Leiden, 376–396.

# Appendix 1

## EXTENDED SUMMARY IN SPANISH

# Procesamiento espectrotemporal auditivo humano en ruido y su relación con la percepción del habla

Resumen extendido en español

Tesis Doctoral

**David López Ramos**

Directores

Enrique A. López Poveda

Luis Enrique López Bascuas

Septiembre 2024

Instituto de Neurociencias de Castilla y León

Universidad de Salamanca

# RESUMEN

El sistema auditivo se adapta al entorno acústico para asegurar una representación fiable y eficiente de las características sonoras perceptualmente relevantes. Mecanismos fisiológicos como el reflejo olivococlear medial (MOCR), el reflejo del músculo del oído medio (MEMR) o la adaptación sináptica neural pueden ser activados por ruido y cambiar la codificación neural del sonido en el tiempo. Por lo tanto, la selectividad frecuencial auditiva, la sensibilidad a la modulación y la percepción del habla pueden cambiar dinámicamente en entornos ruidosos. El objetivo general de esta tesis fue investigar los posibles cambios inducidos por el ruido en la selectividad frecuencial y la sensibilidad a la modulación, y su posible relación con la adaptación al ruido en el reconocimiento de palabras.

La tesis consta de tres estudios. El primer estudio exploró si la duración del estímulo afecta las estimaciones de la sintonización coclear humana en frecuencias bajas y altas. La sintonización coclear se estimó utilizando un método de enmascaramiento hacia adelante con ruido horadado (NN). Los niveles de la máscara en el umbral de enmascaramiento se midieron para duraciones de máscaras de 30 y 400 ms. La hipótesis era que las estimaciones de sintonización podrían ser diferentes cuando las máscaras son lo suficientemente largas como para activar el MEMR y/o el MOCR. La duración de la máscara no tuvo efecto en la sintonización a 4 kHz. En contraste, el uso de máscaras cortas produjo estimaciones de sintonización coclear más amplias a 500 Hz.

El segundo estudio investigó el efecto de los precursores de ruido en las curvas de sintonización psicoacústica humana (PTCs) a 500 Hz y 4 kHz. Los estímulos utilizados para medir las PTCs eran demasiado cortos (máscaras de 30 ms y sondas de 10 ms) para activar el MEMR o el MOCR por sí mismos. Los precursores de ruido (ruido de banda ancha de 300 ms a 60 dB SPL) fueron diseñados para activar el MOCR con una mínima activación del MEMR. Los precursores contralaterales no tuvieron un efecto significativo en las PTCs. Los precursores de ruido ipsilaterales y bilaterales estrecharon las PTCs a 500 Hz pero las ensancharon a 4 kHz. Se utilizó un modelo computacional, inspirado fisiológicamente, de enmascaramiento anterógrado con control eferente para simular e interpretar los resultados. Las PTCs simuladas fueron ampliamente consistentes con las PTCs experimentales y sugirieron que, aunque el patrón de resultados sea aparentemente diferente en las dos frecuencias, es consistente con la inhibición de la ganancia del amplificador coclear por activación del MOCR mediante los precursores.

El tercer estudio investigó si la adaptación al ruido ocurre en la detección de modulación espectral (SM), temporal (TM) y espectrotemporal (STM) así como en el reconocimiento del habla. El reconocimiento de palabras vocodizadas, y la sensibilidad a TM y SM en ruido mejoraron cuando la palabra o la señal modulada se retrasaron 800 ms desde el inicio del ruido. En contraste, esta mejora no ocurrió en el reconocimiento de palabras naturales o en la detección de STM. En conjunto, estos hallazgos sugieren que la adaptación al ruido en el reconocimiento del habla probablemente no esté mediada por mejoras en la codificación de STM.

En conjunto, los hallazgos muestran que la selectividad de frecuencia auditiva humana y la sensibilidad a las modulaciones espectrales y temporales pueden cambiar en presencia de ruido. Esto tiene implicaciones tanto para la comprensión de la selectividad de frecuencia auditiva humana como para el diseño de métodos y estímulos óptimos para estimarla. Además, estos resultados destacan la importancia de incorporar mecanismos de adaptación al ruido en los modelos de la periferia auditiva para simular con precisión el procesamiento auditivo espectrotemporal en entornos ruidosos.

**Palabras clave:** adaptación al ruido, adaptación del rango dinámico, selectividad frecuencial auditiva, sintonización coclear, reflejo olivococlear medial, reflejo del músculo del oído medio, percepción del habla en ruido, percepción de modulaciones espectrotemporales.

# ABREVIATURAS Y ACRÓNIMOS

Algunos de los acrónimos empleados en esta tesis se han tomado directamente del inglés.

AM: amplitud modulada.
AN: nervio auditivo (del inglés *auditory nerve*)
ANOVA: Análisis de la varianza (del inglés *analysis of variance*).
BM: membrana basilar (del inglés *basilar membrane*).
CAS: estimulación acústica contralateral (del inglés contralateral acoustic stimulation)
c/o: ciclos por octava.
dB: decibelio.
DRA: adaptación del rango dinámico (del inglés *dynamic range adaptation*).
ERB: ancho de banda rectangular equivalente, (del inglés *equivalent rectangular bandwidth*).
FM: frecuencia modulada
Hz: Hercio.
HSD: diferencia honestamente significativa (del inglés *honestly significant difference*).
IHC: célula ciliada interna (del inglés *inner hair cell*).
MEMR: reflejo del músculo del oído medio (del inglés *middle-ear muscle reflex*).
MOCR: reflejo olivococlear medial (del inglés *medial olivocochlear reflex*).
ms: milisegundo.
NN: ruido horadado (del inglés *notched noise*).
NNTC: curva de sintonía obtenida con el método de ruido horadado (del inglés *notched-noise tuning curve*).
OHC: célula ciliada externa (del inglés *outer hair cell*).
PTC: curva de sintonía psicoacústica (del inglés *psychoacoustical tuning curve*).
Roex: exponencial redondeada (del inglés *rounded-exponential*).
SFOAE: emisión otoacústica de estímulo frecuencial (del inglés *stimulus-frequency otoacoustic emission*)
SL: sensación sonora (del inglés *sensation level*).
SM: modulación espectral (del inglés *spectral modulation*).
SPL: nivel de presión sonora (del inglés *sound pressure level*).
SRT: umbral de recepción verbal (del inglés *speech reception threshold*).
STM: modulación espectrotemporal (del inglés *spectro-temporal modulation*).
TFS: estructura fina temporal (del inglés *temporal fine structure*).
TM: modulación temporal (del inglés *temporal modulation*).

# 1.

# INTRODUCCIÓN GENERAL

## 1.1. MOTIVACIÓN

El sistema auditivo codifica y procesa las propiedades físicas del sonido que son relevantes para la percepción. Este proceso se realiza de manera dinámica, ya que el sistema auditivo se adapta a determinados aspectos del entorno acústico para asegurar una representación fiable y eficiente de las características espectrotemporales de los sonidos (ver, por ejemplo, Willmore y King, 2023). Pese a esto, algunos aspectos importantes de la percepción auditiva como la selectividad frecuencial, la sensibilidad a las modulaciones espectrales (SM), temporales (TM) y espectrotemporales (STMs), o su relación con la percepción del habla, a menudo se estudian como si fueran estáticos y no sujetos a procesos de adaptación.

Varios mecanismos fisiológicos, como el reflejo olivococlear medial (MOCR), el reflejo del músculo del oído medio (MEMR) o la adaptación sináptica neuronal, podrían cambiar la forma en que los sonidos se codifican en el sistema auditivo a lo largo del tiempo en entornos ruidosos. Estos mecanismos pueden hacer que la selectividad frecuencial, la sensibilidad a las modulaciones y la percepción del habla cambien dinámicamente en entornos ruidosos. Estudios psicoacústicos han demostrado que la selectividad frecuencial humana puede cambiar en presencia de ruidos ipsilaterales (Strickland, 2001; Jennings et al., 2009; Jennings y Strickland, 2012b) y contralaterales (Kawase et al., 2000; Quaranta et al., 2005; Vinay y Moore, 2008; Aguilar et al., 2013; Wicher, 2013; Wicher y Moore, 2014). También se ha demostrado que la detección de tonos con amplitud modulada (AM) en ruido (Almishaal et al., 2017; Marrufo-Pérez et al., 2018b; Wojtczak et al., 2019) y el reconocimiento del habla en ruido (Ben-David et al., 2012,2016; Marrufo-Pérez et al., 2018a) pueden mejorar a medida que las señales se retrasan en el ruido.

No obstante, la investigación existente sobre la adaptación al ruido es limitada y muchos aspectos sobre los efectos de la adaptación en la percepción auditiva siguen siendo desconocidos o están todavía por esclarecer. Por ejemplo, se desconoce si la adaptación al ruido afecta la selectividad frecuencial de la misma manera en frecuencias altas y bajas o para ruidos de distintas lateralidades con respecto al oído de prueba. Además, no está claro si los efectos de la adaptación al ruido en la selectividad frecuencial son consistentes con alguno de los mecanismos fisiológicos mencionados anteriormente. En cuanto a la sensibilidad a las modulaciones, aún no se ha investigado si la sensibilidad a SM, TM y STM mejora cuando las señales moduladas se retrasan en el ruido. Por último, aunque está bien establecido que el reconocimiento del habla mejora a medida que el habla se retrasa en el ruido, no está claro si esta mejora está relacionada con una representación mejorada de la envolvente temporal o

espectral del habla, como se sugiere en estudios anteriores (Ainsworth y Meyer, 1994; Marrufo-Pérez et al., 2018a,2020).

El objetivo general de esta tesis es caracterizar los posibles cambios inducidos por el ruido en la selectividad frecuencial y la sensibilidad a las modulaciones, y su posible relación con la adaptación al ruido en el reconocimiento de palabras. Específicamente, la tesis investiga los cambios en la selectividad frecuencial y en la sensibilidad a las modulaciones en diferentes condiciones de ruido. También investiga la posible relación entre la adaptación al ruido en la sensibilidad a SM, TM y STM y la adaptación en el reconocimiento de palabras.

## 1.2. ANTECEDENTES

### *1.2.1. Adaptación al ruido*

El sistema auditivo se adapta constantemente al entorno acústico (revisado por Willmore y King, 2023). Esta adaptación tiende a mejorar la función auditiva a medida que los sonidos de interés se retrasan en el ruido de fondo, un fenómeno originalmente denominado "overshoot" (Elliot, 1965; Zwicker, 1965a, b) pero más recientemente conocido como adaptación al ruido (Cervera y Ainsworth, 2005; Cervera y González-Álvarez, 2007; revisado en Marrufo-Pérez y López-Poveda, 2022).

La adaptación al ruido probablemente resulta de un desplazamiento inducido por el ruido en el rango dinámico de las neuronas auditivas hacia el nivel sonoro de ruido de fondo (Evans, 1975) y/o hacia el nivel sonoro más común en el entorno (Dean et al., 2005). En un entorno ruidoso, las neuronas auditivas ajustan su rango dinámico para minimizar la saturación de su tasa de descarga y, por lo tanto, mejorar la codificación tanto de la envolvente temporal como espectral de los estímulos (**Fig. 1.1A**, **1.1B**). No está claro si la adaptación del rango dinámico (DRA) hacia el nivel sonoro del ruido de fondo y la DRA hacia el nivel sonoro más común en el entorno son procesos fundamentalmente diferentes o esencialmente el mismo. La DRA hacia el nivel sonoro de fondo se ha estudiado midiendo las respuestas neuronales a tonos cortos en ruido de fondo continuo. En esta situación, las neuronas auditivas disminuyen su tasa máxima de disparo y desplazan su rango dinámico hacia el nivel de ruido. Este fenómeno ocurre a lo largo de todo el sistema auditivo (nervio auditivo (AN): Evans, 1975; Costalupes et al., 1984; Gibson et al., 1985; núcleos cocleares: Gibson et al., 1985; May y Sachs, 1992; colículo inferior: Rees y Palmer, 1989; corteza auditiva: Phillips y Hall, 1986). En contraste, la DRA hacia el nivel más frecuente en el entorno se ha estudiado midiendo las funciones tasa de descarga-nivel sonoro para neuronas auditivas en respuesta a tonos que cambian dinámicamente en nivel sonoro según una distribución preestablecida de niveles, en la que un conjunto de niveles se presenta el 80% del tiempo, constituyendo una "región de alta probabilidad" (Dean et al., 2005). En esta situación, las neuronas auditivas adaptan su rango dinámico hacia los niveles en la región de alta probabilidad. Esta forma de DRA también se ha encontrado a lo largo de todo el sistema auditivo (AN: Wen et al., 2009, 2012; colículo inferior: Dean et al., 2005, 2008; corteza auditiva: Watkins y Barbour, 2008, 2011).

La adaptación del rango dinámico tarda entre 100 y 400 ms en ocurrir (Wen et al., 2012). Se han sugerido varios mecanismos fisiológicos con un curso temporal aproximadamente similar

como posibles causantes de la DRA, incluyendo el MEMR (Costalupes et al., 1984; Gibson et al., 1985; Grange et al., 2022), el MOCR (Dean et al., 2005, 2008; Wen et al., 2009, 2012; Zilany y Carney, 2010; Grange et al., 2022) y procesos sinápticos relacionados con la exocitosis y endocitosis de vesículas de neurotransmisores entre las células ciliadas internas (IHCs) y las fibras AN (Dean et al., 2005, 2008; Wen et al., 2009, 2012; Zilany y Carney, 2010).

La activación del MEMR reduce la amplitud de la vibración del estribo transmitida a la ventana oval (Hung y Dallos, 1972). El MEMR tarda entre 100 y 450 ms en activarse (Borg, 1982; Qiu y Stucker, 1998) y puede ser desencadenado por sonidos con niveles superiores a unos 55 dB SPL (Feeney et al., 2017). Dado que el MEMR puede atenuar el nivel de fondo y su curso temporal es aproximadamente como el de la DRA, esto sugiere que el MEMR podría estar involucrado en la DRA. Sin embargo, datos no publicados discutidos por Gibson et al. (1985) mostraron que la DRA también ocurre después de cortar los tendones del músculo del oído medio.

Las neuronas olivococleares mediales se proyectan desde las regiones mediales del complejo olivar superior hasta las células ciliadas externas (OHCs) en la cóclea (Warr y Guinan, 1979; revisado en Lopez-Poveda, 2018). La activación de las neuronas olivococleares mediales reduce la ganancia del amplificador coclear a niveles sonoros bajos y moderados, y así linealiza las respuestas de la membrana basilar (BM) (Murugasu y Russell, 1996; Dolan et al., 1997; Cooper y Guinan, 2003, 2006). En entornos ruidosos, el ruido podría linealizar la respuesta de la BM mediante la activación de las vías olivococleares mediales ipsilaterales y/o contralaterales. Esta linealización podría mejorar la representación de la envolvente acústica en la salida de la BM (**Fig. 1.1C**). Además, la activación del MOCR descomprime y desplaza la función tasa-nivel de las fibras AN a niveles más altos (Winslow y Sachs, 1987, 1988). Hay algunas evidencias que apoyan el papel del MOCR en la adaptación al ruido. Por ejemplo, el MOCR presenta diferentes constantes de tiempo de activación, pero, en general, la constante de tiempo global de la activación del MOCR es de alrededor de 300 ms (Backus y Guinan, 2006), lo que es consistente con el curso temporal de la DRA en las fibras AN. Además, cuando el haz olivococlear se corta, las fibras AN no muestran DRA (Kawase et al., 1993). Sin embargo, se ha encontrado que la adaptación al ruido en el reconocimiento del habla (Marrufo-Pérez et al., 2018a) y la detección de AM (Marrufo-Pérez et al., 2019a) ocurre en los usuarios de implantes cocleares, para quienes los efectos del MOCR están ausentes porque la estimulación eléctrica proporcionada por el implante omite las OHCs (Lopez-Poveda et al., 2016).

Procesos fisiológicos en las sinapsis entre las IHCs y las fibras del AN, o entre las neuronas auditivas centrales, también podrían contribuir a la DRA. Por ejemplo, la liberación de neurotransmisores ocurre en una amplia gama de escalas de tiempo. El componente cinético rápido de la exocitosis de neurotransmisores en la sinapsis IHC-AN que subyace a la adaptación a corto plazo de la AN es probablemente demasiado rápido para contribuir a la DRA, pero los componentes cinéticos más lentos (cuyo curso temporal es de varios cientos de milisegundos) son candidatos prometedores para causar la DRA (Dean et al., 2005, 2008; Wen et al., 2009, 2012; Zilany y Carney, 2010). La desensibilización de los receptores postsinápticos también podría contribuir a la DRA (Wen et al., 2009, 2012; Zilany y Carney, 2010).

Es poco probable que procesos inhibitorios como la supresión mecánica coclear (Cooper, 1996) o la modulación de las fibras AN mediante la activación de las neuronas olivococleares laterales (LOC) estén involucrados en la DRA ya que su curso temporal es demasiado corto (casi instantáneo para la supresión, Ruggero et al., 1992; Tavartkiladze et al., 1996) o demasiado largo [del orden de minutos para la activación de las LOC, Dean et al. (2008), Wen et al. (2009) y Zilany y Carney (2010)] en comparación con el curso temporal de la DRA.

La adaptación al ruido se ha medido en varias tareas auditivas como la detección de tonos puros (también conocida como "overshoot") (Zwicker, 1965a, b; Elliott, 1965; Carlyon y White, 1992; Bacon y Savel, 2004; Fletcher et al., 2013), en la detección de AM superpuesta a un tono puro (Almishaal et al., 2017; Marrufo-Pérez et al., 2018b; Wojtczak et al., 2019), en la detección de modulación de frecuencia (Mesik y Wojtczak, 2020), y en el reconocimiento del habla (Cervera y González-Álvarez, 2007; Ben-David et al., 2012; Marrufo-Pérez et al., 2018a) (revisado por Marrufo-Pérez y López-Poveda (2022)). Sin embargo, la adaptación al ruido en otras áreas importantes de la percepción auditiva, como la selectividad frecuencial, la detección de SM y STM, o la relación entre la adaptación al ruido en esos aspectos y la adaptación en el reconocimiento del habla sigue siendo poco comprendida.

### *1.2.2. Selectividad frecuencial*

La selectividad frecuencial auditiva es la capacidad de percibir por separado los múltiples componentes de frecuencia que constituyen un sonido complejo. Esta capacidad refleja en parte la sintonización de frecuencia de las fibras del AN (Fletcher, 1940; Evans y Wilson, 1973; Evans, 2001), que está en gran medida determinada por la sintonización de las respuestas mecánicas cocleares (Narayan et al., 1998; aunque ver Recio-Spinoso y Oghalai, 2017).

La cóclea actúa como un analizador de frecuencia. Los tonos puros producen un patrón de vibración a lo largo de la BM, conocido como la onda viajera, cuyo pico ocurre en diferentes lugares a lo largo de la BM dependiendo de la frecuencia del tono: los tonos de alta frecuencia producen respuestas máximas cerca de la base de la cóclea, mientras que los tonos de baja frecuencia alcanzan su pico cerca del ápice (Békésy, 1960). La sintonización coclear está estrechamente vinculada al amplificador coclear (Rhode, 1971; revisado en Davis, 1983), un proceso mediante el cual las OHC amplifican la vibración de la BM y agudizan su respuesta a frecuencias específicas. La amplificación coclear también está vinculada a la compresión, el fenómeno por el cual una amplia gama de presiones acústicas se codifica (comprime) en una gama más estrecha de respuestas mecánicas (y neuronales) cocleares (Ruggero y Rich, 1991).

El enmascaramiento auditivo y la selectividad frecuencial son procesos complementarios. En efecto, si un sonido de una determinada frecuencia (la máscara) enmascara a otro sonido (la sonda o señal) de una frecuencia diferente, esto significa que el oyente no puede percibir por separado ambas frecuencias. Debido a esto, el enmascaramiento se utiliza a menudo para investigar la compresión coclear y la selectividad frecuencial en humanos (por ejemplo; Patterson, 1976; Nelson y Freyman, 1984; Glasberg y Moore, 1990; Oxenham y Shera, 2003; Eustaquio-Martín y Lopez-Poveda, 2011; Jennings y Strickland, 2012a). La compresión se ha investigado utilizando curvas de enmascaramiento temporal (Nelson et al. 2001; Lopez-Poveda et al., 2003) y curvas de crecimiento del enmascaramiento (Oxenham y Plack, 1997).

El uso del enmascaramiento para estimar la selectividad frecuencial se ha enmarcado en el modelo de enmascaramiento basado en el espectro de potencia (Fletcher, 1940). Este modelo asume que el sistema auditivo opera como un banco de filtros, cada uno sintonizado a una frecuencia diferente, cuyas anchuras de banda reflejan la capacidad de resolución de frecuencia del sistema auditivo (Fletcher, 1940, Glasberg y Moore, 1990). Según el modelo de enmascaramiento basado en el espectro de potencia, en una tarea de enmascaramiento la señal se detecta utilizando el filtro auditivo que proporciona la mejor relación señal/máscara en su salida, y el umbral de enmascaramiento ocurre cuando la relación señal/máscara en la salida del filtro supera un valor crítico. El filtro en cuestión es el que está centrado en o alrededor de la frecuencia de la señal.

Dos técnicas de enmascaramiento ampliamente usadas para estimar la selectividad frecuencial son las curvas de sintonización psicoacústicas (PTCs) y las curvas de sintonización estimadas con ruido horadado (NNTCs). Una PTC es una representación gráfica del nivel sonoro de la máscara (umbral de enmascaramiento) requerido para enmascarar una señal (típicamente un tono puro de nivel fijo al que se denomina sonda) en función de la frecuencia de la máscara (Small, 1959). La máscara puede ser un tono puro o un ruido de banda estrecha. Para las PTCs medidas en enmascaramiento simultáneo, el uso de tonos puros como máscaras puede dar lugar a la aparición de "pulsaciones" lo que ayudaría espuriamente a detectar la sonda (Kluk y Moore, 2004). Por esta razón suele recomendarse el uso de bandas estrechas de ruido como máscaras. Además, cuando la máscara y la sonda se presentan simultáneamente, la respuesta al tono de sonda puede ser suprimida por la máscara (Sachs y Kiang, 1968; Cooper, 1996). Justamente, para evitar los efectos contaminadores de la supresión auditiva, se recomienda el uso de procedimientos de enmascaramiento anterógrado, esto es, procedimientos en los que la máscara precede en el tiempo a la sonda. Este procedimiento permite una estimación de la selectividad frecuencial que refleja más fielmente la sintonización subyacente de las respuestas de la BM (Shera et al., 2002).

A veces, el filtro que proporciona la mayor relación señal/máscara no es el que está centrado en la frecuencia de la señal. Cuando eso ocurre, la señal podría ser detectada a través de filtros adyacentes con mejores relaciones señal/máscara, un fenómeno conocido como escucha fuera de frecuencia. Para minimizar el efecto que la escucha fuera de frecuencia pueda tener en las estimaciones de selectividad frecuencial, se emplea el método de ruido horadado (NN) (Patterson, 1976). El método NN minimiza la escucha fuera de frecuencia utilizando dos bandas de ruido como máscara, una situada por encima y la otra por debajo de la frecuencia de la señal. Esto fuerza al oyente a depender de la región de frecuencia alrededor de la señal, proporcionando una estimación más precisa del filtro auditivo sintonizado en la frecuencia de la señal. La selectividad frecuencial se infiere de los umbrales de enmascaramiento para una serie de NN con diferentes anchuras entre estas bandas de ruido de alta y baja frecuencia. Una NNTC es la representación gráfica del umbral de enmascaramiento en función de la anchura entre las dos bandas de ruido. Las NNTCs suelen ajustarse asumiendo formas de filtro exponencial redondeado (roex), y la anchura de banda (por ejemplo, el ancho de banda rectangular equivalente (ERB)) del filtro roex ajustado se toma como una estimación de la selectividad frecuencial.

Usando el método NN, estudios anteriores han investigado cómo cambia la selectividad frecuencial humana en función de la frecuencia (Glasberg y Moore, 1990; Oxenham y Shera, 2003; Leschke et al., 2022) (**Fig. 1.2**). Las estimaciones de Glasberg y Moore (1990) se obtuvieron utilizando el método NN aplicado en enmascaramiento simultáneo, con máscaras de nivel fijo, y variando el nivel de la sonda para estimar el umbral de la sonda; es decir, utilizando un enfoque iso-estímulo (Eustaquio-Martín y Lopez-Poveda, 2011). Este enfoque difiere de los enfoques fisiológicos en los que la selectividad frecuencial coclear se mide variando el nivel del estímulo para lograr una respuesta fija (enfoque iso-respuesta). Para hacer que las estimaciones conductuales de la selectividad frecuencial sean lo más comparables posible a las medidas fisiológicas de la sintonización coclear, Shera et al. (2002) aplicaron el método NN en enmascaramiento anterógrado, con sondas de nivel fijo, y variando el nivel de la máscara para medir el umbral enmascaramiento. Las estimaciones obtenidas con este último método fueron más estrechas que las obtenidas por Glasberg y Moore (1990) y sugirieron que la sintonización coclear era más aguda en humanos que en otros mamíferos (ver también Shera et al. (2010), para evidencia de emisiones otoacústicas). Recientemente, los resultados de Oxenham y Shera (2003) se han extendido a frecuencias por debajo de 1000 Hz (Leschke et al., 2022).

Los estudios recién descritos y las estimaciones conductuales más recientes de la selectividad frecuencial humana se obtuvieron sin considerar los posibles efectos de la adaptación descritos en la **Sección 1.2.1**. La adaptación podría alterar la selectividad frecuencial en entornos ruidosos. Estudios previos han sugerido que la selectividad frecuencial humana podría verse afectada por la estimulación acústica contralateral (CAS) (Kawase et al., 2000; Quaranta et al., 2005; Vinay y Moore, 2008; Aguilar et al., 2013; Wicher, 2013; Wicher y Moore, 2014) o por precursores sonoros ipsilaterales (Strickland, 2001; Jennings et al., 2009; Jennings y Strickland, 2012b). Además, existe evidencia psicoacústica [revisada en Lopez-Poveda (2018) y Jennings (2021)] que sugiere que el uso de precursores sonoros puede inhibir la ganancia del amplificador coclear y linealizar las respuestas de la BM (Krull y Strickland, 2008; Jennings et al., 2009; Roverud y Strickland, 2010; Yasin et al., 2014; DeRoy Milvae y Strickland, 2018,2021; Salloom y Strickland, 2021; Salloom et al., 2023).

Aunque esta evidencia sugiere que la selectividad frecuencial puede variar dinámicamente en entornos ruidosos, el tema no se ha investigado sistemáticamente y las diferencias metodológicas entre estudios complican el análisis de aspectos importantes, dejando varias preguntas cruciales sin respuesta. Por ejemplo, las máscaras utilizadas para medir la selectividad frecuencial son a menudo lo suficientemente largas y su nivel lo suficientemente alto como para que pudieran activar algunos de los mecanismos fisiológicos involucrados en la adaptación al ruido. Sin embargo, no está claro si este es el caso o hasta qué punto las estimaciones conductuales de la selectividad frecuencial varían según la duración de la máscara. Además, no está claro si la selectividad frecuencial en ruido es diferente de la selectividad frecuencial en silencio, si depende de la lateralidad del ruido, o si el ruido afecta a la selectividad frecuencial de manera diferente en frecuencias altas y bajas. También, sigue sin estar claro si los posibles efectos del ruido en la selectividad frecuencial son consistentes con los procesos adaptativos discutidos anteriormente, como el MOCR. Uno de los objetivos de la presente tesis es investigar algunos de estos temas.

### *1.2.3. Percepción de modulaciones espectrales y temporales*

Otro aspecto importante de la percepción auditiva es la sensibilidad a SM, TM y STM. Las SM se refieren a variaciones en la energía a lo largo de la frecuencia, mientras que las TM se manifiestan como cambios en la amplitud a lo largo del tiempo. Las STM combinan modulaciones espectrales y temporales simultáneas, apareciendo así en un espectrograma como desplazamientos diagonales de energía (**Fig. 1.3**).

Existen varios métodos para evaluar la sensibilidad a SM, TM y STM. Estos métodos generalmente implican el uso de ruido o complejos multitonales con envolventes espectrales y temporales moduladas. La modulación de las envolventes espectrales y temporales puede caracterizarse por la densidad espectral (en ciclos por octava, c/o) y la tasa temporal de modulación (en Hz), respectivamente, junto con la fase de la modulación. En una tarea de detección de STM, los oyentes deben diferenciar entre un ruido con envolventes espectrales y/o temporales moduladas y un ruido con envolventes espectrales y temporales planas. En esta tarea, la variable dependiente es la profundidad de modulación (Bernstein et al., 2013, 2016; Mehraei et al., 2014; Venn-Davies et al., 2015; Isarangura et al., 2019; Magits et al., 2019; Zaar et al., 2023). En contraste, en una tarea de discriminación de STM, se pide a los participantes que distingan entre ruido modulado espectrotemporalmente y ruido que también está modulado espectrotemporalmente, pero con una densidad espectral diferente (Aronoff y Landsberger, 2013); o una modulación de fase inversa (Supin et al., 1998; Henry et al., 2005; Venn-Davies et al., 2015); o con una dirección STM diferente (Archer-Boyd et al., 2018). En estos casos, la profundidad de modulación permanece fija y la variable dependiente es la densidad de modulación espectral.

Dado que la detección de SM y STM requiere resolver picos y valles espectrales, es razonable pensar que la sensibilidad a SM y STM está limitada por la capacidad de resolución frecuencial del sistema auditivo (Summers y Leek, 1994; Leek y Summers, 1996; Isarangura et al., 2019). A altas densidades de picos y valles espectrales, la detección de SM podría estar limitada por la selectividad frecuencial. Además, el ensanchamiento de los filtros auditivos periféricos encontrado en oyentes con pérdida auditiva (Glasberg y Moore, 1986; Yonemoto et al., 1998) se asocia con una pobre representación interna del perfil espectral (Leek et al., 1987; Summers y Leek, 1994; Leek y Summers, 1996). Combinando datos experimentales y de modelado, Ponsot et al. (2021) mostraron que la diferencia entre oyentes con audición normal y oyentes con pérdida auditiva en el procesamiento espectrotemporal podría explicarse principalmente por los filtros cocleares más anchos que muestran los oyentes con pérdida auditiva. Existe, asimismo, evidencia que sugieren la existencia de una relación entre la detección de SM y STM y la selectividad frecuencial. Por ejemplo, Venn-Davies et al. (2015) encontraron correlaciones significativas entre los umbrales de detección de SM a diferentes densidades espectrales y los ERBs de los filtros auditivos a 500 y 2000 Hz, lo que sugiere que el rendimiento en tareas de selectividad frecuencial y detección de SM podría estar gobernado por mecanismos comunes. De manera similar, Bernstein et al. (2013) encontraron que la detección de STM estaba fuertemente correlacionada con la sensibilidad a la frecuencia modulada (FM) de baja frecuencia y los ERBs de los filtros auditivos centrados en 4 kHz. Mehraei et al. (2014) también encontraron que cuando una STM de 4 c/o y 4 Hz se superponía a una banda de ruido de una octava centrada en 4 kHz, los oyentes con pérdida auditiva tenían umbrales de detección

significativamente peores que los normoyentes, lo cual es consistente con la reducción de la selectividad frecuencial que afecta la sensibilidad a STM.

La sensibilidad a SM, TM y STM se ha estudiado extensamente, por ejemplo, calculando las funciones de transferencia de SM (Summers y Leek, 1994; Eddins y Bero, 2007; Isarangura et al., 2019), TM (Viemeister, 1979) y STM (Chi et al., 1999), que relacionan la sensibilidad a la modulación con la densidad de modulación espectral, la tasa de modulación temporal o ambas simultáneamente. La detección de TM superpuesta a ruidos de banda ancha muestra una tendencia paso-bajo con una sensibilidad a modulaciones temporales relativamente constante hasta 16 Hz y que empeora gradualmente para tasas de modulación más altas (Viemeister, 1979). La sensibilidad a SM alcanza su máximo alrededor de 1 y 4 c/o con una disminución de la sensibilidad para densidades espectrales más bajas y más altas (Summers y Leek, 1994; Eddins y Bero, 2007; Isarangura et al., 2019). Los datos de Chi et al. (1999) muestran que la función de transferencia de STM puede explicarse considerando la sensibilidad a SM y TM como dos variables independientes. Sin embargo, esto ha sido desafiado por evidencia fisiológica (por ejemplo, Escabí y Schreiner, 2002).

Los estudios recién descritos y sus conclusiones se realizaron sin considerar los procesos de adaptación dinámica que pueden afectar la detección de la modulación. Se ha demostrado que para tonos AM en ruido, la detección de AM mejora a medida que los tonos se retrasan en el ruido (Marrufo-Pérez et al., 2018b) o son precedidos por precursores de ruido (Almishaal et al., 2017; Wojtczak et al., 2019) (revisado en Marrufo-Pérez y Lopez-Poveda, 2022). Esto sugiere que la sensibilidad a TM puede ser mejorada por procesos de adaptación que ocurren en el sistema auditivo. Sin embargo, se sabe poco sobre cómo la sensibilidad a SM, TM y STM se ve afectada por el ruido de fondo, o si los mecanismos de adaptación al ruido mencionados anteriormente pueden mejorar la sensibilidad a la modulación. Otro objetivo de la presente tesis es investigar la adaptación al ruido en la detección de SM, TM y STM.

### *1.2.4. Percepción del habla en ruido*

La adaptación al ruido en el reconocimiento del habla se ha estudiado de varias maneras. Un método implica variar el retraso de inicio del ruido del habla y medir la mejora (si la hay) en el porcentaje de palabras reconocidas (Ainsworth y Meyer, 1994; Cervera y Ainsworth, 2005; Cervera y González-Álvarez, 2007) o en el umbral de recepción del habla (SRT) (es decir, la relación señal/ruido al 50% de reconocimiento de palabras) (Ben-David et al., 2012, 2016; Marrufo-Pérez et al., 2018a) a medida que los estímulos del habla se retrasan en el ruido. Otro método implica el uso de precursores de ruido. En este caso, la adaptación se calcula como la diferencia entre el porcentaje de palabras reconocidas o los SRT con y sin el precursor de ruido (Marrufo-Pérez et al., 2020). Investigaciones previas han mostrado que la adaptación al ruido puede mejorar el reconocimiento de palabras hasta en un 30% (Ainsworth y Meyer, 1994; Cervera y Ainsworth, 2005; Cervera y González-Álvarez, 2007) y los SRT hasta en 5 dB (Ben-David et al., 2012, 2016; Marrufo-Pérez et al., 2018a). Sin embargo, aún no está claro a qué se deben estas mejoras.

El reconocimiento del habla depende de una codificación precisa de las SM, TM y STM contenidas en el habla (**Fig. 1.3**) (Elliot y Theunissen, 2009). El reconocimiento empeora cuando

el habla se degrada espectralmente (Ainsworth y Millar, 1972; van Veen y Houtgast, 1985; Zahorian y Jagharghi, 1993; Liu y Eddins, 2008) o temporalmente (Drullman et al., 1994), aunque no es posible reducir la información temporal en el habla sin afectar la información espectral o viceversa (Drullman et al., 1996). Además, hay evidencia de que el reconocimiento del habla empeora cuando se eliminan las modulaciones espectrales bajas (<4 ciclos/kHz) y las modulaciones temporales lentas (<12 Hz) del habla. (por ejemplo, Elliot y Theunissen, 2009).

Varios estudios han investigado la relación entre el reconocimiento del habla y la sensibilidad a SM (Henry et al., 2005; Venn-Davies et al., 2015) y STM (Bernstein et al., 2013, 2016; Mehraei et al., 2014; Zaar et al., 2023). Los umbrales de detección de SM (Venn-Davies et al., 2015) y los umbrales de discriminación de SM (Henry et al., 2005; Venn-Davies et al., 2015) son predictores significativos del reconocimiento del habla en ruido. La relación entre STM y la inteligibilidad del habla depende de la densidad espectral, la tasa de modulación temporal y la medida de inteligibilidad del habla. Sin embargo, en general, la sensibilidad a STM explica una parte significativa de la variancia de la inteligibilidad del habla que no se explica por medidas de audibilidad como el audiograma o el índice de inteligibilidad del habla (Bernstein et al., 2013, 2016; Mehraei et al., 2014; Zaar et al., 2023). Esto sugiere que la mejora en el reconocimiento de palabras reportada en estudios previos de adaptación al ruido puede deberse a una mejor codificación de las envolventes espectrales y/o temporales del habla en el sistema auditivo.

Estudios previos han proporcionado evidencia en apoyo de esta hipótesis. Por ejemplo, el hecho de que la detección de tonos con AM en ruido mejore con el tiempo sugiere que la adaptación al ruido en el reconocimiento del habla ocurre debido a una mejora en la codificación de la envolvente temporal del habla. Esto se apoya aún más en el hecho de que la adaptación al ruido es mayor para palabras procesadas para mantener solo las señales de la envolvente temporal (es decir, palabras vocodizadas) que para palabras no procesadas que mantienen las señales de la envolvente y la estructura temporal fina (TFS) (Marrufo-Pérez et al., 2018a, 2020). Un estudio de modelado computacional (Ainsworth y Meyer, 1994) mostró que cuando se introduce un desplazamiento del umbral hacia el nivel de ruido en la función tasa de respuesta-nivel sonoro de las fibras del AN del modelo, la envolvente espectral del habla se codifica mejor en el AN. Cuando la salida del modelo del AN se envió a un reconocedor automático de habla, el reconocimiento automático del habla aumentó debido a esta mejora en la codificación de la envolvente espectral del habla. Hasta la fecha, sin embargo, ningún estudio ha investigado si la adaptación al ruido en el reconocimiento de palabras está asociada con una mejor codificación de las señales espectrales o temporales del habla. Otro objetivo de la presente tesis es investigar esta posibilidad.

## 1.3. OBJETIVOS E HIPÓTESIS

El objetivo principal de esta tesis es investigar experimentalmente el efecto del ruido de fondo en el procesamiento auditivo espectrotemporal humano y si los efectos inducidos por el ruido en el procesamiento espectrotemporal están relacionados con la adaptación al ruido en el reconocimiento del habla. La hipótesis general es que la selectividad frecuencial y la sensibilidad a la modulación cambian en entornos ruidosos (debido a mecanismos de

adaptación fisiológica presentes en el sistema auditivo). Este cambio en el procesamiento espectrotemporal podría estar relacionado con la adaptación al ruido en el reconocimiento de palabras.

Los objetivos específicos y la hipótesis son los siguientes:

1. Investigar si las estimaciones conductuales más recientes de la sintonización coclear humana (Shera et al., 2002; Leschke et al., 2023) se ven afectadas (confundidas) por la adaptación al ruido debido al uso de máscaras largos y de alto nivel sonoro. Se plantea la hipótesis de que las máscaras de larga duración podrían activar el MOCR y el MEMR y, por lo tanto, afectar las estimaciones conductuales de la sintonización coclear.
2. Investigar el potencial efecto de los precursores de ruido contralaterales, ipsilaterales y bilaterales en las estimaciones de selectividad frecuencial. Se plantea la hipótesis de que los precursores de ruido podrían afectar el ancho de banda del filtro auditivo de una manera consistente con la activación del MOCR por los precursores. También se explora si el efecto del precursor de ruido en la selectividad frecuencial difiere en frecuencias bajas y altas.
3. Investigar si la sensibilidad a SM, TM y STM en ruido mejora a medida que las señales se retrasan desde el inicio del ruido. Como se explicó anteriormente (**Fig. 1.1**), la DRA neuronal puede mejorar la codificación de las envolventes moduladas espectral y temporalmente, por lo que se plantea la hipótesis de que la sensibilidad a SM, TM y STM mejora cuando SM, TM y STM se retrasan en el ruido.
4. Investigar si (y en qué medida) la adaptación al ruido en el reconocimiento de palabras ocurre debido a una mejora en la sensibilidad a SM, TM y/o STM. Se plantea la hipótesis de que los mismos oyentes que mostraron adaptación al ruido en el reconocimiento de palabras también mostrarán adaptación en la detección de modulación presente en los tres tipos de modulación.

## 1.4. CONTRIBUCIONES ORIGINALES

Las principales contribuciones originales de esta tesis son:

1. Evidencia empírica de que las estimaciones conductuales de la sintonización coclear humana a bajas frecuencias, pero no a altas frecuencias, dependen de la duración del estímulo.
2. Evidencia experimental de que los precursores de ruido afectan las estimaciones conductuales de la selectividad frecuencial de manera diferente dependiendo de la lateralidad del ruido y la frecuencia de la señal. Además, evidencia proveniente del modelado computacional de que estos resultados experimentales pueden explicarse por la activación del MOCR por los precursores de ruido.
3. Evidencia empírica de que la sensibilidad a TM superpuesta a un complejo multitonal mejora cuando el portador se retrasa en el ruido.
4. Evidencia empírica de que la sensibilidad a SM, pero no a STM, mejora cuando el portador modulado se retrasa en el ruido.

5. Evidencia experimental de que, en oyentes con audición normal, la adaptación al ruido en la percepción del habla no está relacionada con una mejora en la detección de modulaciones STM en ruido.

## 1.5. DESCRIPCIÓN GENERAL DE LA TESIS

Esta tesis está organizada de acuerdo con los objetivos enumerados en la **Sección 1.3.**

El **Capítulo 2** describe un estudio experimental destinado a investigar el efecto de la duración de la máscara en las estimaciones conductuales de la sintonización coclear humana a 500 Hz y 4 kHz obtenidas utilizando el método de enmascaramiento anterógrado y con el método de NN.

El **Capítulo 3** presenta un estudio destinado a investigar el efecto del precursor de ruido ipsilateral, contralateral y bilateral en las PTCs de 500 Hz y 4 kHz medidas utilizando máscaras de corta duración. Además, se utiliza un modelo computacional inspirado fisiológicamente de enmascaramiento anterógrado con control eferente para simular el efecto de los precursores en las PTCs, asumiendo que los precursores reducen la ganancia del amplificador coclear mediante la activación del MOCR.

El **Capítulo 4** describe un estudio destinado a investigar si la adaptación al ruido ocurre en la detección de modulaciones espectrales, temporales y espectrotemporales. Para los mismos participantes (diseño de medidas repetidas), también se mide la adaptación al ruido en el reconocimiento de palabras bajo las mismas condiciones para investigar si la adaptación en el reconocimiento del habla está relacionada con mejoras en la codificación de SM, TM y/o STM.

El **Capítulo 5** ofrece una discusión integral y unificada de los hallazgos principales, destacando su importancia y discutiendo sus implicaciones y limitaciones. Además, este capítulo sugiere nuevos desafíos y describe posibles direcciones experimentales futuras.

## 1.6. UNIDADES

Todas las unidades siguen las convenciones del Sistema Internacional de Unidades.

## 1.7. ÉTICA

El trabajo reportado en esta tesis implicó la realización de experimentos conductuales en sujetos humanos y la recopilación de datos personales. Los procedimientos fueron aprobados por la Junta de Revisión Ética de la Universidad de Salamanca (Salamanca, España). Todos los participantes fueron voluntarios adultos y no recibieron pago por su tiempo. Firmaron un consentimiento informado antes de ser admitidos en los experimentos.

# 2.

# EFECTO DE LA DURACIÓN DE LA MÁSCARA EN ESTIMACIONES COMPORTAMENTALES DE LA SINTONIZACIÓN COCLEAR HUMANA[1]

## 2.1. INTRODUCCIÓN

Los métodos de enmascaramiento auditivo empleados originalmente para evaluar la selectividad frecuencial conductual han evolucionado a lo largo de los años para inferir la sintonización coclear. Los umbrales de enmascaramiento obtenidos con enmascaramiento anterógrado para máscaras NN y una señal con un nivel sonoro fijo y bajo proporcionan estimaciones precisas de la sintonización coclear (Shera et al., 2002). La sintonización coclear puede variar, por ejemplo, dependiendo del estado de la activación del MOCR (Cooper y Guinan, 2006; revisado en Lopez-Poveda, 2018 y Jennings, 2021). Además, la sintonización coclear varía con el nivel sonoro (en la base coclear: Ruggero et al., 1997; en el ápice: Recio-Spinoso y Oghalai, 2017), y el nivel efectivo que llega al oído interno puede variar dependiendo del estado de activación del MEMR. Las máscaras NN típicamente utilizadas en el método de Shera et al. son lo suficientemente largas y su nivel general es lo suficientemente alto como para que puedan activar uno o ambos reflejos. Por lo tanto, las estimaciones de sintonización inferidas con el método de Shera et al. (2002) usando máscaras largas podrían verse afectadas por el MOCR y/o el MEMR. El objetivo del presente estudio fue investigar si la sintonización coclear humana inferida con el método de Shera et al. (2002) depende de la duración de la máscara. Las formas de los filtros auditivos y la sintonización se inferían usando máscaras cortas (30 ms) y largas (400 ms). Se asumió que las máscaras de 30 ms eran demasiado cortas para activar el MOCR o el MEMR, mientras que las máscaras de 400 ms eran lo suficientemente largas para activar los dos reflejos siempre que su nivel fuera lo suficientemente alto.

[1] El trabajo descrito en este capítulo está presentado en la siguiente publicación: López-Ramos, D., Eustaquio-Martín, A., López-Bascuas, L.E. and Lopez-Poveda, E.A. (in press). Effect of stimulus duration on estimates of human cochlear tuning. *Hearing Research.* Submitted 20 March 2024. Accepted 10 July 2024. https://doi.org/10.1016/j.heares.2024.109080

## 2.2. MATERIAL Y MÉTODOS

Las sondas fueron sinusoides de 20 ms a 10 dB de sensación sonora (SL). Las máscaras fueron ruidos con una muesca espectral colocada simétrica y asimétricamente alrededor de la frecuencia de la sonda. Se midieron los niveles de máscara en el umbral de enmascaramiento en enmascaramiento anterógrado para varias anchuras de NN y para duraciones de máscara de 30 y 400 ms. Siete personas normoyentes participaron en el experimento. Las NNTCs se ajustaron asumiendo formas de filtro roex y el modelo de enmascaramiento basado en el espectro de potencia, y se calcularon los ERBs a partir de los ajustes.

## 2.3. RESULTADOS

A 500 Hz, las NNTCs fueron más planas para la máscara de 30 ms que para la de 400 ms (**Fig. 2.1**), y los ERBs fueron significativamente más amplias para la máscara más corta (**Fig. 2.3**) ($ERB_{30ms}$=126 Hz vs $ERB_{400ms}$=55 Hz, test de Wilcoxon, Z=28, p=0.016). A 4 kHz, por el contrario, los umbrales de máscara fueron más altos para las máscaras más cortas (**Fig. 2.2**), pero los ERBs no fueron significativamente diferentes para las dos duraciones de máscara (**Fig. 2.4**) ($ERB_{30ms}$=294 Hz vs $ERB_{400ms}$=277 Hz, $F_{(1,6)}$=0.56, p=0.482).

## 2.4. DISCUSIÓN

A 4 kHz el método produce estimaciones de sintonización iguales utilizando máscaras cortas (30 ms) y largas (400 ms). Esto indica que las estimaciones de sintonización coclear obtenidas con este método a 4 kHz probablemente no se vean afectadas por el MOCR, MEMR u otros factores de confusión.

Sin embargo, a 500 Hz, el método sugiere una sintonización significativamente más amplia utilizando los enmascaradores más cortos. Diversos factores podrían contribuir a este efecto: por ejemplo, el uso de máscaras largas podría hacer que las estimaciones se vean afectadas por el MEMR o el MOCR. Sin embargo, también el uso de máscaras cortas podría aumentar las posibles confusiones entre máscara y señal, además de aumentar la dispersión espectral. Este resultado sugiere que las estimaciones de sintonización a 500 Hz (ya sean conductuales o fisiológicas) pueden no estar libres de factores contaminantes. Desafortunadamente, puede ser difícil optimizar los estímulos para controlar todos los factores de confusión, ya que las medidas que pueden minimizar el impacto de un factor podrían aumentar el impacto de otro. Por ejemplo, el uso de enmascaradores más largos para minimizar dispersión espectral o la confusión entre máscara y señal aumentaría la posibilidad de que las estimaciones de sintonización se vean afectadas por el MEMR o el MOCR. En otras palabras, a frecuencias bajas, el estímulo utilizado para medir el ancho de banda podría cambiar el propio ancho de banda y/o afectar la medición. Por lo tanto, las estimaciones de sintonización coclear humana a frecuencias bajas deben interpretarse con precaución.

## 2.5. CONCLUSIONES

Los principales resultados se pueden resumir de la siguiente manera:

1) A 500 Hz, los ERBs fueron significativamente más anchos para la máscara más corta ($ERB_{30ms}$=126 Hz vs $ERB_{400ms}$=55 Hz).
2) A 4000 Hz, por el contrario, los ERBs no fueron significativamente distintas cuando se infirieron utilizando máscaras cortas (30 ms) y largas (400 ms) ($ERB_{30ms}$=294 Hz vs $ERB_{400ms}$=277 Hz)
3) Se debe tener precaución al interpretar la selectividad frecuencial conductual en términos de sintonización mecánica coclear en frecuencias bajas.

# 3.

# EFECTOS DE RUIDOS PRECURSORES IPSILATERALES, CONTRALATERALES Y BILATERALES EN CURVAS DE SINTONIZACIÓN PSICOACÚSTICAS HUMANAS[2]

## 3.1. INTRODUCCIÓN

Se piensa que la sintonización coclear y, por lo tanto, la selectividad frecuencial auditiva cambia en entornos ruidosos mediante la activación del MOCR. En humanos, la selectividad frecuencial auditiva suele estimarse midiendo PTCs. El objetivo del presente estudio fue investigar el efecto de la activación del MOCR en la selectividad frecuencial auditiva humana a 500 Hz y 4 kHz cuando el MOCR es activado por ruido de banda ancha ipsilateral, contralateral y bilateral. Para ayudar a la interpretación de los resultados, se utilizó un modelo computacional de enmascaramiento anterógrado inspirado fisiológicamente con control de ganancia coclear eferente (Aguilar et al., 2013) capaz de simular los efectos de los precursores de ruido en las PTCs.

## 3.2. MATERIAL Y MÉTODOS

Las PTCs para señales de 500 Hz y 4 kHz se midieron en enmascaramiento anterógrado utilizando máscaras (30 ms) y señal (10 ms) cortas para minimizar la activación del MOCR por las máscaras o la señal. La señal fue presentada a 10 dB SL, y el nivel de la máscara se varío de manera adaptativa para calcular el umbral de enmascaramiento. Para la sonda de 4 kHz, las frecuencias de la máscara fueron 2.0, 2.8, 3.6, 3.8, 4.0, 4.2, 4.4 y 4.8 kHz; para la sonda de 500 Hz, las frecuencias de la máscara fueron 250, 350, 450, 475, 500, 525, 550, 600, 700 y 800 Hz. El intervalo entre el fin de la máscara y el inicio de la señal fue de 2 ms. Las PTC también se

[2] El trabajo descrito en este capítulo está presentado en la siguiente publicación: López-Ramos, D., López-Bascuas, L.E., Eustaquio-Martín, A. and Lopez-Poveda, E.A. (under review). Effects of ipsilateral, contralateral, and bilateral noise precursors on psychoacoustical tuning curves in humans. *Hearing Research*. Submitted 13 June 2024.

midieron en presencia de precursores de ruido de banda ancha largos (300 ms) a 60 dB SPL ipsilaterales, contralaterales y bilaterales para investigar el efecto del MOCR ipsilateral, contralateral y bilateral en la sintonización de las PTC (**Fig. 3.1**). Cuatro oyentes con audición normal participaron en los experimentos. Cada lado de una PTC fue ajustado de manera independiente a una roex(*pwt*) y se calcularon los ERB a partir de los ajustes.

## 3.3. RESULTADOS

A 500 Hz, los precursores ipsilaterales y bilaterales estrecharon las PTC al disminuir los umbrales para las máscaras con frecuencias cercanas a la frecuencia de la señal con efectos mínimos en los umbrales para las máscaras alejadas en frecuencia de la señal (**Fig. 3.2** y **Fig. 3.4B**). El ANOVA de un factor reveló un efecto significativo del tipo de precursor sobre el ERB a 500 Hz ($F_{(3,9)}$=5.69, p=0.018). La prueba post-hoc de Tukey HSD mostró que el ERB en la condición de control (139 Hz) difirió significativamente del ERB estimada con precursores ipsilaterales (85 Hz, p=0.002) y bilaterales (75 Hz, p<0.001). En contraste, el ERB en la condición de control no difirió del ERB estimado con un precursor contralateral (118 Hz, p=0.49) A 4 kHz, por el contrario, los precursores ipsilaterales y bilaterales apenas afectaron los umbrales para las máscaras cerca de la frecuencia de la señal, pero ampliaron las PTC al reducir los umbrales para las máscaras lejanas a la frecuencia de la señal (**Fig. 3.3** y **Fig. 3.4B**). Los precursores contralaterales apenas afectaron las PTC. El ANOVA de un factor reveló un efecto significativo del tipo de precursor sobre el ERB a 4 kHz ($F_{(3,9)}$=7.24, p=0.009). El post-hoc Tukey HSD mostró que el ERB en la condición de control (210 Hz) difirió significativamente del ERB estimado con precursores ipsilaterales (394 Hz, p<0.001) y bilaterales (392 Hz, p<0.001). Por el contrario, el ERB en la condición de control no difirió del ERB estimado con precursor contralateral (267 Hz, p=0.52).

Se utilizó un modelo computacional con control de ganancia coclear eferente para interpretar los resultados. El modelo sugirió que, a pesar de las aparentes diferencias, el patrón de resultados es consistente con la inhibición de la ganancia coclear por parte del MOCR ipsilateral y bilateral de manera similar en las dos frecuencias de sonda y más fuertemente que el MOCR contralateral (**Fig. 3.4C** y **D** y **Fig. 3.5**).

## 3.4. DISCUSIÓN

En línea con estos resultados, algunos estudios han encontrado que los cambios en la selectividad frecuencial coclear causados por la activación del MOCR también pueden ser diferentes entre frecuencias. Un estudio que midió las curvas de sintonización de supresión de emisiones otoacústicas de estímulo frecuencial (SFOAEs) encontró una sintonización más estrecha a 1 kHz, pero más ancha a 2 y 4 kHz cuando el MOCR fue activado por CAS (Xing y Gong, 2017). El ensanchamiento encontrado aquí a 4 kHz es consistente con mediciones previas de selectividad frecuancial en presencia de precursores ipsilaterales (Strickland, 2001; Jennings et al., 2009; Jennings y Strickland, 2012b) y CAS (Quaranta et al., 2005; Vinay y Moore, 2008). Hasta donde sabemos, ningún estudio ha medido aún los efectos de precursores ipsilaterales en PTCs de baja frecuencia. En el **Capítulo 2**, se mostró que el ancho de banda del

filtro auditivo de 500 Hz es menor cuando se estima usando enmascaradores NN más largos (400 ms) que más cortos (30 ms). En el presente estudio, también encontramos que las PTCs de 500 Hz son más estrechas en presencia de un precursor de ruido ipsilateral diseñado para activar el MOCR. Esto apoya la hipótesis de que el MOCR puede estar involucrado en el efecto de la duración del enmascarador sobre las estimaciones conductuales de la sintonización coclear reportadas en el **Capítulo 2**.

Los precursores utilizados fueron diseñados (en nivel y duración) para activar el MOCR con una mínima activación del MEMR (Lilaonitkul y Guinan, 2009a, 2009b; Aguilar et al., 2013) y es improbable que las máscaras y señales utilizados activaran el MOCR debido a su corta duración. Por lo tanto, se asumió que los efectos del precursor se deben a la inhibición de la ganancia coclear causada por la activación del MOCR. Las simulaciones del modelo sugieren que el estrechamiento de las PTCs de 500 Hz y el ensanchamiento de las PTCs de 4 kHz en presencia de precursores ipsilaterales y bilaterales pueden surgir de una interacción de la inhibición de ganancia eferente (causada por la activación del MOCR por los precursores) con la compresión que afecta a las máscaras (**Fig. 3.6**). Sin embargo, los resultados del modelo deben interpretarse con cautela. Por ejemplo, Aguilar et al. (2013) mostraron que el CAS a 60 dB SPL no activó el MEMR. Por lo tanto, aquí se ha asumido que los precursores de ruido utilizados (a 60 dB SPL) no activaron el MEMR. Aunque esto podría ser cierto para los precursores contralaterales, el MEMR se activa más fácilmente por sonidos ipsilaterales que por contralaterales (McGregor et al., 2018). Por lo tanto, no podemos descartar que los precursores solos (o combinados con las máscaras) activaran el MEMR.

## 3.5. CONCLUSIONES

En este estudio se investigó los efectos de los precursores de ruido ipsilaterales, contralaterales y bilaterales en las PTCs medidas con máscaras cortas. Los principales hallazgos fueron:

1) Los precursores contralaterales apenas afectaron las PTCs a 500 Hz o 4 kHz.
2) A 500 Hz, los precursores ipsilaterales y bilaterales estrecharon las PTCs al disminuir el umbral para las máscaras alrededor del pico de las PTCs.
3) A 4 kHz, los precursores ipsilaterales y bilaterales ensancharon las PTCs al disminuir el umbral para las frecuencias de máscaras en las colas de las PTCs.
4) El patrón de resultados es aproximadamente consistente con la inhibición de la ganancia coclear causada por la activación del MOCR por los precursores.

# 4.

# ADAPTACIÓN AL RUIDO EN LA DETECCIÓN DE MODULACIONES ESPECTROTEMPORALES Y EN EL RECONOCIMIENTO DE PALABRAS[3]

## 4.1. INTRODUCCIÓN

La adaptación al ruido es la mejora en la función auditiva a medida que la señal de interés se retrasa en el ruido. Con respecto al reconocimiento del habla, los oyentes con audición normal pueden reconocer hasta un 30% más de sílabas y tolerar 3 dB más de ruido mientras mantienen un 50% de reconocimiento correcto de palabras cuando la señal del habla se retrasa ~300 ms desde el inicio del ruido (Cervera y Gonzalez-Alvarez, 2007; Ben-David et al., 2012; Marrufo-Pérez et al., 2018a). Aún no está claro si la adaptación al ruido en el reconocimiento del habla resulta de una mejora en la codificación de las modulaciones temporales, espectrales o espectrotemporales del habla. Uno de los objetivos del presente estudio fue investigar experimentalmente si existe adaptación al ruido en la detección de modulaciones espectrales y espectrotemporales. Otro objetivo fue investigar qué tipo de señal de modulación puede estar involucrada en la adaptación al ruido en el reconocimiento del habla. La hipótesis era que, si la adaptación ocurre para los tres tipos de modulaciones (temporal, espectral y espectrotemporal), entonces la adaptación en el reconocimiento de palabras podría estar relacionada con mejoras en la codificación de una o más de esas tres señales de modulación. Sin embargo, si la adaptación no ocurre para la detección de modulación espectral o espectrotemporal, entonces el tipo de modulación en cuestión probablemente no estaría involucrado en la adaptación al ruido en el reconocimiento del habla.

[3] El contenido de este capítulo ha sido adaptado del siguiente artículo: López-Ramos, D., Marrufo-Pérez, M.I., Eustaquio-Martín, A., López-Bascuas, L.E. and Lopez-Poveda, E.A. (in press). Adaptation to noise in spectrotemporal modulation detection and word recognition. *Trends in Hearing*. Submitted 27 November 2023. Accepted 12 June 2024. https://doi.org/10.1177/23312165241266322

## 4.2. MATERIAL Y MÉTODOS

Dieciocho adultos con audición normal participaron en los experimentos. En las tareas de detección de modulación, la señal era un ruido ondulado modulado espectral y/o temporalmente de 200 ms. La densidad de modulación espectral fue de 2 c/o, la tasa de modulación temporal fue de 10 Hz y las modulaciones espectrotemporales combinaron estas dos modulaciones, lo que resulta en una modulación espectrotemporal descendente. Se realizó un experimento de control para determinar si los resultados se generalizaban a modulaciones espectrotemporales ascendentes. En la tarea de reconocimiento del habla, la señal consistió en palabras bisilábicas naturales o vocodizadas. Se midieron los umbrales de detección de modulación a 0 dB relación señal-ruido y los umbrales de recepción del habla en silencio y en ruido blanco (a 60 dB SPL) para retrasos de inicio de señal-ruido de 50 ms (condición temprana) y 800 ms (condición tardía) (**Fig. 4.1**). La adaptación se calculó como la diferencia de umbral entre las condiciones temprana y tardía.

## 4.3. RESULTADOS

Los análisis estadísticos mostraron que la mejora del SRT en la condición tardía no fue significativa para palabras naturales (0.6 dB) (prueba t de Student pareada: $t_{(17)}$=1.90, p=0.19), pero fue significativa para palabras vocodizadas (2.1 dB) (prueba t de Student pareada: $t_{(17)}$=5.44, p<0.001) (**Fig. 4.3A**). La adaptación fue significativamente menor para palabras naturales (0.59 dB) que para palabras vocodizadas (2.1 dB) (prueba t de Student pareada: $t_{(17)}$=-3.50, p=0.004) (**Fig. 4.4A**).

La adaptación al ruido fue estadísticamente significativa para la detección de SM (prueba t de Student pareada: $t_{(17)}$= 5.19, p<0.001) y TM (prueba t de Student pareada: $t_{(17)}$=4.37, p=0.002), pero no para la detección de STM (prueba t de Student pareada: $t_{(17)}$=-0.19, p=1.0). (**Fig. 4.3B**). La adaptación al ruido en la detección de SM (2.1 dB) y TM (2.2 dB) fue significativamente diferente de la adaptación en la detección de STM (-0.04 dB) (prueba t de Student pareada: $t_{(17)}$=4.56, p=0.001; prueba t de Student pareada: $t_{(17)}$=4.15, p=0.003, respectivamente). Pero la diferencia entre la adaptación al ruido en la detección de SM y TM no fue significativa (prueba t de Student pareada: $t_{(17)}$=-0.21, p=1.0) (**Fig. 4.4B**).

La adaptación en la detección STM ascendente (-0.4 ± 1.1 dB) no fue estadísticamente significativa (**Fig. 4.5A**) (prueba t de Student pareada: $t_{(17)}$=-1.32, p=0.894) y no difirió de la adaptación en la detección de modulaciones STM descendente (prueba t de Student: $t_{(29)}$= 1, p= 0.327) (**Fig. 4.5B**).

## 4.4. DISCUSIÓN

Este es el primer estudio que demuestra empíricamente una mejora en la detección de SM a medida que la señal de SM se retrasa desde el inicio del ruido. Este resultado es consistente con las simulaciones computacionales de Ainsworth y Meyer (1994). Con respecto a la detección de TM, el mecanismo para detectar ondulaciones temporales superpuestas en un ruido podría ser el mismo que se utiliza para detectar AM en tonos puros.

Desafortunadamente, una comparación directa entre la adaptación al ruido en la detección de TM, tal como se midió aquí, y la adaptación encontrada en estudios previos de detección de AM no es sencilla porque la cantidad de adaptación depende de la tasa de modulación (Viemeister, 1979; Sheft y Yost, 1990; Almishaal et al., 2017), el nivel sonoro (Almishaal et al., 2017; aunque ver Wojtczak et al., 2019), la lateralidad y el tipo de ruido (Marrufo-Pérez et al., 2018b; Wojtczak et al., 2019), el tipo de estímulo [tono puro (Marrufo-Pérez et al., 2018b; Wojtczak et al., 2019) vs. ruido (Viemeister, 1979; Sheft y Yost, 1990; Almishaal et al., 2017)], y el paradigma de prueba [uso de precursores (Almishaal et al., 2017; Wojtczak et al., 2019) vs. retraso desde el inicio del ruido (Marrufo-Pérez et al., 2018b)]. Para una revisión de la adaptación al ruido en la detección de AM, ver Marrufo-Pérez y López-Poveda (2022). La razón de la falta de adaptación al ruido en la detección de STM es desconocida, pero se hipotetiza que esto podría deberse a procesos de integración espectral y temporal en centros superiores del sistema auditivo (**Fig. 4.6**).

Encontramos que la adaptación al ruido ocurre en la detección de SM y TM, así como en el reconocimiento de palabras vocodizadas. Esto indica que la adaptación al ruido en el reconocimiento de palabras (vocodizadas) puede estar mediada por mejoras en la codificación neuronal de SM y/o TM. Sin embargo, la adaptación no ocurrió en la detección de STM. Esto indica que es poco probable que la adaptación al ruido en el reconocimiento de palabras (vocodizadas) esté mediada por mejoras en la codificación neuronal de señales de habla de STM.

## 4.5. CONCLUSIONES

1. Para señales en ruido, la sensibilidad a las modulaciones espectrales y temporales, pero no a las modulaciones espectrotemporales, mejora a medida que la señal se retrasa 800 ms desde el inicio del ruido. En otras palabras, la adaptación al ruido ocurre en la detección de modulaciones espectrales y temporales, pero no en la detección de modulaciones espectrotemporales combinadas.
2. El reconocimiento de palabras en ruido puede mejorar a medida que las palabras se retrasan 800 ms en el ruido. La mejora fue estadísticamente significativa para las palabras vocodizadas, pero no para las palabras naturales.
3. Por lo tanto, la adaptación al ruido en el reconocimiento de palabras puede estar mediada por mejoras en la codificación neuronal de modulaciones espectrales o temporales sostenidas, pero es poco probable que esté mediada por mejoras en la codificación de modulaciones espectrotemporales combinadas.

# 5.

# DISCUSIÓN GENERAL

En entornos ruidosos, las neuronas auditivas pueden ajustar sus rangos dinámicos hacia el nivel de ruido y hacia el nivel más común en el entorno (por ejemplo, Evans, 1975; Phillips y Hall, 1986; Dean et al., 2005). Varios mecanismos fisiológicos (incluyendo el MEMR, el MOCR y/o procesos sinápticos) pueden causar DRA. La adaptación del rango dinámico puede evitar que la tasa de disparo de la neurona se sature y, por lo tanto, mejorar la codificación de las envolventes temporales y espectrales acústicas (**Fig. 1.1**). Sin embargo, frecuentemente se asume que la capacidad del sistema auditivo para codificar y procesar información espectral o temporal, es estática y no se ve afectada por procesos de adaptación. Uno de los objetivos de esta tesis fue explorar empíricamente si la selectividad frecuencial auditiva y la sensibilidad a las SM, TM y STM se ven afectadas por la adaptación al ruido. Otro objetivo fue investigar si la adaptación al ruido en el reconocimiento del habla (i.e. la mejora en el reconocimiento del habla cuando el habla se retrasa desde el inicio del ruido) se relaciona con una mejora en la codificación de las envolventes temporales o espectrales del habla, como se propone en estudios previos (Ainsworth y Meyer, 1994; Marrufo-Pérez y Lopez-Poveda, 2022).

Se han realizado dos estudios para investigar el posible efecto de los procesos de adaptación en la selectividad frecuencial auditiva humana. El primer estudio (descrito en el **Capítulo 2**) exploró si la duración del estímulo afecta las estimaciones de sintonización coclear humana en frecuencias bajas y altas. La sintonización coclear se estimó utilizando el método de Shera et al. (2002), una técnica de enmascaramiento conductual diseñada para proporcionar estimaciones de selectividad frecuencial lo más cercanas posible a la sintonización coclear. La hipótesis era que las estimaciones de sintonización podrían ser diferentes cuando máscaras NN utilizadas en el método fueran lo suficientemente largas como para activar el MEMR y/o el MOCR. Se ha demostrado que la duración de la máscara tiene un efecto insignificante en las estimaciones de sintonización a 4 kHz. En contraste, el uso de máscaras cortas produce estimaciones más anchas de la sintonización a 500 Hz.

En el segundo estudio (descrito en el **Capítulo 3**) se investigó el efecto de los precursores de ruido en las PTCs a 500 Hz y 4 kHz. Los estímulos utilizados para medir las PTCs (tanto la sonda como las máscaras) tenían una duración breve que hacía improbable la activación del MEMR o el MOCR. Sin embargo, los precursores de ruido fueron diseñados para activar el MOCR con una mínima activación del MEMR. Por consiguiente, dado el diseño experimental, parece razonable asumir que cualquier efecto de los precursores en las PTCs procede de la activación del MOCR. Los resultados experimentales pusieron de manifiesto que los precursores contralaterales tuvieron un efecto insignificante en las PTCs. Sin embargo, los precursores ipsilaterales y bilaterales estrecharon las PTCs a 500 Hz, pero ensancharon las PTC a 4 kHz. Es

decir, los precursores de ruido ipsilaterales y bilaterales afectaron las PTC de manera diferente en frecuencias bajas y altas. Se utilizó un modelo computacional inspirado fisiológicamente de enmascaramiento anterógrado con control eferente (utilizado en Aguilar et al., 2013) para simular e interpretar estos resultados en relación con una posible activación del MOCR. Las PTCs simuladas fueron razonablemente consistentes con las PTCs experimentales. Lo que sugiere que, aunque el patrón de resultados es aparentemente diferente en las dos frecuencias, los resultados son consistentes con una inhibición de la ganancia del amplificador coclear mediante la activación del MOCR por los precursores de ruido.

Por último, se realizó un estudio experimental para investigar si la sensibilidad a SM, TM, y STM mejora a medida que la señal modulada se retrasa en el ruido (**Capítulo 4**). Se ha demostrado que este es realmente el caso para la detección de SM y TM, pero no para la detección de STM. Además, se ha demostrado que, para los mismos sujetos, la adaptación al ruido se produjo en el reconocimiento de palabras vocodizadas, pero no naturales. En conjunto, estos hallazgos sugieren que la adaptación al ruido en el reconocimiento del habla probablemente no esté relacionado a mejoras en la codificación de STM.

## 5.1. IMPLICACIONES

El resultado de que las estimaciones conductuales de la sintonización coclear humana pueden variar dependiendo de la duración del estímulo, al menos a 500 Hz, tiene implicaciones para nuestra comprensión de la selectividad frecuencial auditiva humana y para el diseño de métodos y estímulos óptimos para estimarla. El método de enmascaramiento anterógrado con máscaras NN empleado por Shera et al., 2002 es una técnica conductual bien controlada para inferir la sintonización coclear. Estudios previos han demostrado que produce estimaciones de sintonización que se ajustan razonablemente bien a las inferidas a partir de la latencia de SFOAEs (en humanos: Shera et al., 2002), así como con mediciones directas de respuestas del AN (en hurones: Sumner et al., 2018). Sin embargo, la evidencia de una correspondencia entre las estimaciones conductuales, SFOAEs y neuronales se limita a frecuencias ≥1 kHz y a duraciones de enmascaradores ≥200 ms. Aún queda por demostrar que las estimaciones conductuales corresponden con los datos de SFOAE y/o neuronales a frecuencias más bajas (<1 kHz) y para máscaras de menor duración.

Aquí, hemos demostrado que a 4 kHz el método de Shera et al. (2002) produce estimaciones de sintonización iguales utilizando máscaras cortas (30 ms) y largas (400 ms). Esto indica que las estimaciones de sintonización coclear obtenidas con este método a 4 kHz probablemente no se vean afectadas por el MOCR, el MEMR u otros factores potencialmente contaminadores discutidos en el **Capítulo 2**. Sin embargo, a 500 Hz, las estimaciones de sintonización son significativamente más ancha utilizando las máscaras más cortas. Como se discutió en el **Capítulo 2**, esta dependencia de la duración de las máscaras podría reflejar una interacción no lineal de mecanismos fisiológicos que pueden afectar la sintonización coclear (por ejemplo, el MEMR y/o MOCR) con factores acústicos (por ejemplo, la dispersión espectral) y cognitivos que pueden afectar el rendimiento humano en la tarea. Puede ser difícil controlar estos factores, ya que las medidas que pueden minimizar el impacto de un factor podrían aumentar el impacto de otro. Por ejemplo, usar máscaras más largas para minimizar la dispersión

espectral aumentaría la posibilidad de que las estimaciones se vean afectadas por el MEMR o el MOCR. En otras palabras, a frecuencias bajas, el estímulo utilizado para estimar la sintonización podría cambiar el propio ancho de banda y/o afectar la medición. Por lo tanto, las estimaciones de la sintonización coclear humana en frecuencias bajas (por ejemplo, Leschke et al., 2022) deben interpretarse con precaución.

El hecho de que la selectividad frecuencial y la sensibilidad a SM y TM cambian en situaciones de ruidosas (**Capítulo 3** y **4**) tiene implicaciones en la comprensión del procesamiento auditivo espectrotemporal humano en ruido. Por ejemplo, sugiere que para simular con precisión el procesamiento auditivo espectrotemporal en entornos ruidosos, es fundamental incorporar mecanismos de adaptación al ruido en los modelos del sistema auditivo periférico. Tradicionalmente, pocos modelos han incluido mecanismos como el MEMR o MOCR para explicar estos cambios inducidos por el ruido en el procesamiento espectrotemporal (ver Lopez-Poveda, 2005, para una revisión completa de los modelos del sistema auditivo periférico). El modelo utilizado en el **Capítulo 3** incorpora un parámetro de atenuación eferente para explicar la inhibición de la ganancia coclear causada por la activación del MOCR. Sin embargo, la inhibición de ganancia en ese modelo no es dinámica ya que el valor del parámetro de atenuación se fijó durante toda la presentación del estímulo. En una versión más reciente del modelo (Grange et al., 2022), la atenuación efferente del MOCR depende dinámicamente de la respuesta del modelo, lo cual es fisiológicamente más realista. Otro modelo periférico que incorpora adaptación (Zilany y Carney, 2010) ha simulado con éxito el fenómeno de DRA en las neuronas del AN al nivel sonoro más frecuente en el entorno (Wen et al., 2009). Este modelo también se ha actualizado para incorporar el control dinámico de la ganancia coclear del MOCR dependiendo de las entradas excitatorias del colículo inferior y el núcleo coclear (Farhadi et al., 2023).

## 5.2. LIMITACIONES Y TRABAJO FUTURO

### *5.2.1. Efecto del MEMR y MOCR en las estimaciones de la selectividad frecuencial*

El presente trabajo tuvo como objetivo investigar los cambios inducidos por el ruido en la selectividad frecuencial auditiva humana. Específicamente, comparamos las estimaciones de la selectividad frecuencial obtenidas usando máscaras NN largas y cortas (**Capítulo 2**), y con y sin precursores de ruido (**Capítulo 3**). La discusión de los resultados se basa en determinados supuestos, a saber, que las máscaras largas y los precursores fueron suficientes para activar el MEMR o el MOCR. Para futuros estudios, valdría la pena evaluar la activación de los dos reflejos directamente utilizando métodos fisiológicos en lugar de hacer suposiciones sobre si los estímulos utilizados activaron esos reflejos. Por ejemplo, el grado de activación del MOCR producido por precursores o máscaras de larga duración podría cuantificarse midiendo la supresión de otoemisiones acústicas (Marrufo-Pérez et al., 2021) y el grado de activación del MEMR podría evaluarse realizando la prueba de contracción del MEMR (Lilaonitkul y Guinan, 2009a,b; Aguilar et al., 2013). Estas mediciones fisiológicas, pueden permitir una evaluación más directa y precisa de la potencial influencia del MEMR y el MOCR en las estimaciones de selectividad frecuencial. También servirían para asegurar que los estímulos utilizados para

medir la selectividad frecuencial no activen los reflejos mientras que los estímulos diseñados para activar los reflejos realmente lo hacen.

### *5.2.2. Relación entre selectividad frecuencial, detección de modulaciones y reconocimiento del habla en ruido*

Como se comentó en el **Capítulo 1**, estudios previos han mostrado una fuerte relación entre la selectividad frecuencial, la sensibilidad a STM y el reconocimiento del habla en ruido (Bernstein et al., 2013; Davies-Venn et al., 2015). Aquí, solo se ha estudiado la relación entre la sensibilidad a SM, TM y STM y el reconocimiento del habla en ruido (**Capítulo 4**). Aunque algunos participantes participaron en todos los experimentos de selectividad frecuencial, sensibilidad a STM y reconocimiento de palabras, el número de esos participantes fue insuficiente para calcular correlaciones entre las tres medidas con suficiente poder estadístico. Además, los cambios inducidos por el ruido en la selectividad frecuencial no se han medido utilizando el mismo paradigma que el cambio inducido por el ruido en la detección de modulación y el reconocimiento de palabras (uso de precursores vs. retraso de la señal con respecto al inicio del ruido). Esta diferencia metodológica impide sacar conclusiones sobre la relación entre estos tres aspectos de la percepción auditiva en ruido. En el futuro, sería interesante ir más allá de la relación bien establecida entre la selectividad frecuencial, detección de modulaciones y el reconocimiento de palabras en ruido, e investigar si los cambios inducidos por el ruido en el ancho de banda de los filtros auditivos están relacionados con las mejoras inducidas por el ruido en la detección de modulaciones y el reconocimiento de palabras. Una hipótesis sería que, si la selectividad frecuencial en ruido mejora como resultado de los mecanismos de adaptación al ruido, esto estaría relacionado con una mejora en la detección de modulaciones en ruido y en el reconocimiento de palabras.

### *5.2.3. Uso de ruido blanco y ruido de banda ancha con espectro plano*

En los experimentos de esta tesis, la adaptación al ruido se ha estudiado utilizando ruido de banda ancha con un espectro plano (**Capítulo 3**) o ruido blanco (**Capítulo 4**). No está claro que el uso de ruidos de banda ancha con espectros planos sea el tipo de ruido más adecuado para estudiar la adaptación al ruido. Por un lado, los estudios de SFOAEs han encontrado que los ruidos de banda ancha son los activadores más eficientes del MOCR (Lilaonitkul y Guinan, 2009a), además, el umbral de activación del MEMR también tiende a ser más bajo con ruidos de banda ancha (Keefe et al., 2010). Por otro lado, en las tareas de detección de modulación y reconocimiento del habla, el ruido modulado y el habla tienen un espectro descendente y el uso de ruidos de espectro plano podría obligar a los participantes a depender de la información de baja frecuencia para detectar la modulación y reconocer el habla. La cantidad de adaptación al ruido en el reconocimiento del habla reportada en el **Capítulo 4** fue similar a la encontrada en estudios previos en los que el espectro del ruido produce la misma cantidad de enmascaramiento en todos los componentes de frecuencia del habla (por ejemplo, utilizando un ruido espectro del habla) (Marrufo-Pérez et al., 2018a; 2020). Pero en el caso de la detección de modulación, no está claro hasta qué punto la adaptación depende de la disponibilidad de esas señales de baja frecuencia. Por esa razón, en futuros experimentos se podrían explorar la adaptación al ruido en la detección de TM y/o SM superpuestas a señales de banda ancha

utilizando ruido de fondo con el mismo espectro que la señal. También sería interesante explorar el efecto de ruidos de fondo más naturales o ecológicos en la adaptación al ruido en el reconocimiento del habla y en la detección de modulaciones para explorar la posible influencia de la adaptación al ruido en entornos de escucha más ecológicos (Marrufo-Pérez et al., en preparación).

### *5.2.4.Adaptación al ruido en la detección de TM y SM*

En el **Capítulo 4**, se evaluó la adaptación al ruido en la detección de SM y TM para una sola combinación de parámetros de densidad espectral y tasa temporal (2 c/o y 10 Hz, respectivamente). Sería interesante explorar en futuras investigaciones cómo varía la adaptación al ruido en TM y SM dependiendo de estos parámetros. Como se comentó en el **Capítulo 1**, la sensibilidad a TM y SM cambia con la tasa de modulación y la densidad de modulación espectral, respectivamente. Específicamente, la función de transferencia de TM muestra una tendencia de paso bajo donde la sensibilidad a TM permanece constante entre 2 y 16 Hz y disminuye gradualmente para tasas de modulación más altas (Viemeister, 1979), mientras que la función de transferencia de SM alcanza su pico alrededor de 1 a 4 c/o (Summers y Leek, 1994; Eddins y Bero, 2007; Isarangura et al., 2019). Investigaciones previas indican una gran mejora en la detección de modulación para frecuencias de modulación ≤ 20 Hz cuando la señal modulada se superpone al final de una banda continua de ruido en comparación a cuando la modulación aparece simultáneamente con el ruido (Viemeister, 1979; Sheft y Yost, 1990). De manera similar, la mejora en la detección de AM cuando es precedida por un precursor disminuye a medida que la tasa de AM aumenta de 20 Hz a 500 Hz (Almishaal et al., 2017). Esto da cierto apoyo a la hipótesis de que la adaptación al ruido puede estar directamente relacionada con la sensibilidad, es decir, cuanto mayor sea la sensibilidad a la AM, mayor será la cantidad de adaptación al ruido en la detección de AM. Aunque esta idea contradice estudios previos, que encontraron una relación inversa entre la adaptación al ruido y el rendimiento en el reconocimiento de palabras (Marrufo-Pérez et al., 2018a, 2020), valdría la pena investigar si la relación se mantiene en la adaptación al ruido en la detección de SM y TM superpuestos a señales de banda ancha.

Otra limitación del presente trabajo es que la adaptación al ruido en la detección de modulación se midió para un nivel de ruido fijo (60 dB SPL) y una relación señal-ruido fija (0 dB). Sería interesante estudiar si la adaptación al ruido en la detección de SM y TM varía con el nivel sonoro y/o relación señal-ruido. Estudios previos han investigado el efecto del nivel sonoro (Wojtczak et al. 2019) y relación señal-ruido (Almishaal et al., 2017; Mesik y Wojtczak, 2020) en la adaptación al ruido en la detección de AM. Sin embargo, aún no se ha investigado si el nivel sonoro y la relación señal-ruido afectan la adaptación cuando TM y SM están superpuestos a estímulos de banda ancha. Estos datos podrían arrojar luz sobre las contribuciones de diferentes mecanismos fisiológicos a la adaptación al ruido medida de manera conductual (Marrufo-Pérez y Lopez-Poveda, 2022). Por ejemplo, si la adaptación al ruido estuviera relacionada con el MEMR, la adaptación debería ser mayor a niveles de ruido altos porque el MEMR es más fuerte a niveles altos. Si la adaptación al ruido estuviera relacionada con el MOCR, entonces la adaptación debería ser mayor a niveles medios donde ocurre la compresión coclear y el MOCR tiene un mayor efecto. Por otro lado, si la adaptación

al ruido estuviera más relacionada con la DRA neuronal hacia el nivel de ruido, entonces la adaptación conductual debería ser mayor a una relación señal-ruido cercana a 0 (Mesik y Wojtczak, 2020).

### *5.2.5.Ausencia de adaptación al ruido en la detección de STM*

Se ha demostrado en el **Capítulo 4** que la adaptación al ruido no ocurre en la detección de STM. Sin embargo, este hallazgo queda restringido a una densidad modulación espectral de 4 c/o y una tasa de modulación temporal de 10 Hz. Podría ser interesante investigar si la falta de adaptación en la detección de STM está limitada a estos parámetros y si la adaptación podría ocurrir para STM con otras ratios entre la tasa y la densidad de modulación, es decir, con otras pendientes en el espectrograma (ver, por ejemplo, Oetjen y Verhey, 2015). Sería razonable esperar que la adaptación en la detección de STM ocurra gradualmente a medida que el STM se vuelve más vertical (por lo tanto, más parecido a TM) o más horizontal (por lo tanto, más parecido a SM). Esto podría arrojar luz sobre la hipótesis (propuesta en el **Capítulo 4**) de que la falta de adaptación al ruido en la detección de STM podría deberse a procesos de integración espectral y temporal en los centros auditivos centrales. Estos procesos de integración contrarrestarían el beneficio de la DRA para la codificación de STMs. A medida que la pendiente del STM se vuelve más pronunciada o más plana en el espectrograma, el efecto perjudicial de la integración espectral y temporal sería menor, y la adaptación al ruido sería mayor.

La falta de adaptación al ruido en la detección de STM también puede indicar que el mecanismo para detectar STM está más estrechamente relacionado con el mecanismo para detectar FM que con el mecanismo para detectar SM o TM. Bernstein et al. (2013) encontraron que la sensibilidad a FM a una frecuencia de modulación de 2 Hz aplicada a una señal de 500 Hz era un predictor significativo de la sensibilidad a STM. Esto sugiere que FM y STM podrían ser detectados por mecanismos similares y, por lo tanto, indica que el uso de *phase-locking* para monitorizar cambios en los picos espectrales puede ser un mecanismo clave para detectar STM. Mientras que se ha demostrado que la DRA mejora la codificación del nivel sonoro en el AN (Evans, 1975; Gibson et al., 1985), no está claro si existe un mecanismo de adaptación que podría mejorar el *phase-locking* en el AN. Si FM y STM fueran detectadas utilizando *phase-locking*, y si mecanismo no mejorara con el tiempo en entornos ruidosos, la detección de FM en ruido, al igual que la detección de STM, no debería mejorar cuando la señal de FM se retrasa desde el inicio del ruido. Esto también sería consistente con el hecho de que la adaptación al ruido en el reconocimiento de palabras naturales (que depende en parte de la codificación de la TFS, Hopkins y Moore (2009)) es menor que la adaptación al ruido en el reconocimiento de palabras vocodizadas.

Estudios previos han demostrado que la detección de FM puede mejorar cuando la señal de FM se retrasa desde el inicio de la señal portadora (Bryne et al., 2013) o cuando es precedida por un precursor (Mesik y Wojtczak, 2020). Sin embargo, la detección de FM en Byrne et al. (2013) y Mesik y Wojtczak (2020) podría deberse a la detección de pistas de AM, un mecanismo conocido como conversión de FM a AM (Zwicker, 1970; Moore y Sek, 1992, 1994, 1995; Whiteford y Oxenham, 2015; Moore et al., 2019; Mesik y Wojtczak, 2020; Whiteford et al., 2020) ya que han utilizado señales portadoras de alta frecuencia moduladas en frecuencia a tasas rápidas. Por esta razón, la mejora en la detección de FM reportada por Byrne et al. (2013)

y Mesik y Wojtczak (2020) probablemente no se deba a una mejor codificación utilizando *phase-locking*. Sería interesante probar experimentalmente si hay adaptación al ruido en la detección de FM cuando FM no puede ser detectada basándose en pistas de AM (por ejemplo, cuando la frecuencia de la señal portadora y la tasa de modulación son bajas, o cuando una señal de FM también está modulada en amplitud para evitar detectar FM utilizando pistas de AM). Si la sensibilidad a STM está relacionada con la detección de FM en tales condiciones, como sugiere Bernstein et al. (2013), esperaríamos una ausencia de adaptación al ruido en la detección de FM, lo que sugeriría, a su vez, que la sensibilidad a STM y la detección de FM, se basan en la codificación de TFS.

### *5.2.6. Ausencia de correlación entre adaptación al ruido en el reconocimiento de palabras y en la detección de modulaciones*

La falta de correlación observada entre la adaptación al ruido en el reconocimiento de palabras y la detección de modulaciones (**Capítulo 4**) podría deberse a múltiples causas que podrían investigarse en futuros estudios. Primero, podría reflejar la limitada variabilidad en nuestra muestra, que solo incluía participantes con audición normal. Estudios previos que reportaron correlaciones significativas entre la sensibilidad a STM y la percepción del habla en ruido (Bernstein et al., 2013, 2016; Zaar et al., 2023) incluyeron en su muestra a oyentes con pérdida auditiva. La inclusión de participantes con pérdida auditiva podría potencialmente generar correlaciones más altas, ya que la muestra tendría mayor variabilidad en los umbrales de detección de modulación, los SRT y la adaptación.

En segundo lugar, también es posible que la mejora en el reconocimiento de palabras se base en otras claves acústicas (o pistas de habla) en lugar de las envolventes espectrales y temporales. Estas otras claves acústicas podrían ser información transitoria, como las transiciones de formantes, que se ha reportado que son esenciales para la discriminación de fonemas (Liberman et al., 1954; Varnet et al., 2013). En este sentido, el uso de técnicas como la correlación inversa (Varnet et al. 2013, 2015; Varnet y Lorenzi, 2022; Osses y Varnet, 2024) podría ayudar a dilucidar, si las señales del habla utilizadas en el reconocimiento del habla difieren a medida que el habla se retrasa desde el inicio del ruido.

En tercer lugar, las tareas de reconocimiento del habla y detección de modulación pueden diferir en la estrategia utilizada por el oyente durante la tarea. En una tarea de detección de modulación de 3 alternativas y respuesta cerrada, se supone que los participantes discriminan entre un estímulo objetivo (modulado) y dos estímulos estándar (no modulados), presentados secuencialmente. Por el contrario, el proceso de reconocimiento del habla comprende la discriminación y selección entre elementos léxicos competidores, que son activados por la presentación acústica de una palabra (Luce y Pisoni, 1998). Suponiendo que la presentación acústica de una palabra activa elementos léxicos similares con modulaciones espectrotemporales similares, los participantes deben discriminar y seleccionar la palabra correcta de un conjunto de candidatos de palabras similares. En base a esto, es posible que se encuentre una relación más fuerte entre la discriminación de modulación y el reconocimiento del habla que entre la detección de modulación y el reconocimiento del habla, ya que tanto la discriminación de modulación como el reconocimiento del habla implican distinguir entre dos o más patrones de modulación, en lugar de solo detectar si un estímulo está modulado. Esto

resultaría en una mayor relación entre la cantidad de adaptación al ruido en ambas tareas. Sin embargo, cabe señalar que Davies-Venn et al. (2015) encontraron valores de correlación similares entre la detección y la discriminación de modulación y el reconocimiento del habla en ruido. Por otro lado, dado que el reconocimiento de palabras también está afectado por procesos léxicos, como el número de palabras léxicamente análogas y la frecuencia de estos elementos léxicos similares (Luce y Pisoni, 1998), es posible que la simple mejora en la codificación de las envolventes temporales y espectrales, que presumiblemente es el factor clave en la mejora de la detección de modulación, se vea desdibujada por la influencia de estos aspectos léxicos, resultando en correlaciones no significativas entre los valores de adaptación al ruido en ambas tareas. Esto sugiere que si el reconocimiento del habla no dependiera de factores léxicos (por ejemplo, reconocimiento de sílabas sin sentido o pseudopalabras), sería posible encontrar una correlación entre la mejora en la representación de SM y TM y la mejora en el reconocimiento de estos elementos lingüísticos.

En cuarto lugar, como se discutió en el **Capítulo 4**, hay otros fenómenos además de la DRA neural que pueden contribuir a la adaptación al ruido. La falta de correlación entre la adaptación al ruido en el reconocimiento de palabras y la detección de modulación puede deberse a la acción de estos factores de confusión. Por ejemplo, es posible que la adaptación al ruido en el reconocimiento de palabras se deba a una mejora atencional, mientras que la adaptación al ruido en la detección de modulación podría deberse a algún tipo de enmascaramiento en el dominio de la modulación (Gallun y Hafter, 2006). Podría ser de interés para futuros estudios examinar específicamente el efecto de estos factores de confusión.

# CONCLUSIONES

Esta tesis tuvo como objetivo investigar si la selectividad frecuencial auditiva humana y la sensibilidad a las modulaciones espectrales, temporales y espectrotemporales se ven afectadas por los procesos de adaptación al ruido. Otro objetivo fue investigar si la adaptación al ruido en el reconocimiento del habla se relaciona con mejoras en la sensibilidad a las señales del habla temporales, espectrales y/o espectrotemporales.

Las principales conclusiones son:

(1) En situaciones ruidosas, la selectividad frecuencial auditiva humana y la sensibilidad a las modulaciones espectrales y temporales cambian a lo largo del tiempo.

(2) La duración de la máscara no afecta las estimaciones comportamentales de la sintonización coclear humana a 4 kHz. Esto sugiere que las estimaciones de sintonización a altas frecuencias no están influenciadas por mecanismos de adaptación al ruido. Por el contrario, el uso de máscaras cortas produce estimaciones más anchas de la sintonización a 500 Hz que cuando se usan máscaras largas. Esto indica que las estimaciones de sintonización coclear a 4 kHz no se ven afectadas por los procesos de adaptación, pero deben interpretarse con precaución a 500 Hz.

(3) Los precursores de ruido contralaterales apenas afectan a la selectividad frecuencial auditiva humana a bajas (500 Hz) o altas frecuencias (4 kHz). Por el contrario, los precursores ipsilaterales y bilaterales aumentan la selectividad frecuencial a bajas frecuencias y la reducen a altas frecuencias. Aunque sean aparentemente diferentes, los dos patrones son ampliamente consistentes con la inhibición eferente de la ganancia coclear por una activación del MOCR por los precursores.

(4) El reconocimiento de palabras en ruido mejora a medida que las palabras se retrasan en el ruido. La sensibilidad a las modulaciones espectrales y temporales, pero no a las modulaciones espectrotemporales, mejora cuando las señales moduladas se retrasan desde el inicio del ruido. Esto sugiere que la adaptación al ruido en el reconocimiento de palabras puede estar relacionada a una mejora en la codificación neuronal de las envolventes espectrales y temporales del habla, pero no a una mejor sensibilidad a las modulaciones espectrotemporales.