sciscan

Study on Hierarchical Attentional Control Mechanisms of Language Comprehension in Noisy Environments
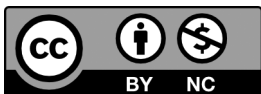
—Exploration of Psycholinguistic Perspectives and Teaching Applications Based on the Cocktail Party Effect

Xuanru Kong

Heilongjiang University, Harbin

Abstract: In the study of language comprehension in noisy environments, the Cocktail Party Effect, as a classic cognitive phenomenon, has vividly revealed the selective attentional capacity of the human auditory system. The effect describes the remarkable phenomenon of an individual's ability to focus on a specific conversation even in the midst of a noisy party, and is essentially a mechanism of selective attention of the attentional control system to key cues of language. Early acoustic processing enhances the efficiency of speech parsing through perceptual preferences developed through everyday language interactions (e.g., voice onset time that reinforces bursts). This cognitive mechanism has a dual significance in noisy language comprehension: on the one hand, it reflects the filtering function underlying attentional control over acoustic interference; on the other hand, its efficiency differences directly reflect the plasticity characterizing the individual's language processing ability. Based on the theoretical framework of psycholinguistics, this study systematically elucidates the mechanism of hierarchical attentional control for language comprehension in noisy environments, using the cocktail party effect as the entry point. Early acoustic processing enhances speech parsing efficiency through experience-driven perceptual weight adjustments (e.g., voice onset time for reinforcement of bursts); mid-phrase lexical competition resists semantic interference through inhibitory control; and late syntactic processing relies on strategic allocation of working memory resources. Behavioral plasticity is characterized by a significant increase in noise comprehension accuracy due to bilinguals' inhibitory control, and children with specific language impairment (SLI) showed significant between-group differences in sentence repetition ability under classroom noise through 8 weeks of rhyming attention training, a finding that provides an empirical basis for language teaching.

Keywords: Cocktail Party Effect; Attentional control; Language comprehension; Noisy environments; Psycholinguistics; Cognitive plasticity; Working memory; Bilingual processing

1 Dynamic Conditioning Mechanisms: A Core Cognitive Basis for Language Comprehension in Noisy Environments

The cocktail party effect is essentially a mechanism of selective attention (Cherry, 1953): locking on to a target linguistic stream across multiple acoustic inputs, a mechanism that forms the dynamic cognitive basis for language comprehension in noisy environments. This modulation has a significant linguistic bias: when background noise overwhelms speech, attention is prioritized to retain articulatory features that are critical for lexical slicing (e.g., the burstiness of consonants or the direction of resonance peaks in vowels (Mattys, 2012)). Attentional distraction leads to reduced efficiency in speech parsing: discrimination accuracy for minimal oppositional words (e.g., "bag/pad") shows significant between-group differences when attention is forced to be distracted (Mattys & Wiget, 2011). This bias suggests that native speakers develop perceptual preferences for specific acoustic cues (e.g., word-initial bursts) through everyday language interactions. Developmental studies have shown that 5-year-olds have significantly higher rates of instruction-following errors than adults in noisy environments, while 10-year-olds no longer perform significantly differently from adults (Escudero et al., 2016).

This phenomenon is consistent with Broadbent's (1958) filtering model. The model suggests that the attentional system acts as an "information gate", allowing only part of the input to enter higher cognitive processing. For example, in a cocktail party scene, listeners prioritize the acoustic features of the target speech (e.g., the pitch of a particular speaker) through the filtering model, while temporarily storing unattended background noise in short-term memory. This selective filtering mechanism explains why people maintain the ability to parse key linguistic cues in noisy environments (Broadbent, 1958).

2 Cognitive Processing Hierarchy Analysis of Attentional Control

Hierarchical attentional control refers to a cascaded three-stage mechanism governing robust language comprehension in noise:

(1) Early acoustic tuning: Dynamic weighting of critical speech cues (e.g., VOT enhancement) to optimize perceptual input;

(2) Mid-stage lexical arbitration: Inhibitory control suppressing competing lexical-semantic activation;

(3) Late syntactic integration: Strategic working memory allocation for complex structure parsing.

This cascading processing comprises three functionally distinct yet interactive stages: early adaptive tuning of acoustic cues provides the base filtering for phonological perception, mid-stage real-time arbitration of lexical competition defends against semantic interference through inhibitory control, and late-stage strategic allocation of syntactic resources relies on the dynamic deployment of working memory. These three stages do not operate in isolation, but form a continuous processing stream through the cascading regulation of the attentional system: the optimal selection of acoustic cues directly influences the efficiency of lexical access (Mattys, 2012), whereas the inhibitory efficacy of lexical competition in turn shapes the resource allocation strategy for syntactic parsing (McMurray et al., 2010). This inter-level interaction is particularly pronounced in noisy environments—when background interference is enhanced, the attentional system prioritizes the reinforcement of early acoustic filtering (e.g., boosting low-frequency

sensitivity in male speech) while compressing the timing of lexical competition through inhibitory control, ultimately reserving more working memory resources for the parsing of complex syntactic structures (Rönnberg et al., 2013). The specific mechanisms of each stage and their synergistic role in noisy language comprehension will be analyzed separately below.

2.1 Dynamic Tuning of Acoustic Cues: Optimization Mechanisms for Speech Perception in Noisy Environments

In noisy environments, speech understanding begins with the extraction of acoustic features, a process that is modulated by the attentional system. The attention system is able to dynamically adjust the perception strategy according to the environmental features. For example, when the background noise is from a female speaker, the system automatically enhances its sensitivity to low-frequency acoustic features of male speech (Mattys et al., 2005) and this adaptive tuning can be directly observed through behavioral experiments: for example, in environments containing mixed male and female noises, the subjects' recognition accuracy of the target male speech was significantly higher than in female-only noise environments (Mattys et al., 2005).

Research in developmental psychology reveals that this cognitive tuning mechanism occurs throughout language development. Children rely more on dynamic acoustic properties such as resonant peak transitions when recognizing speech segments such as fricatives, whereas adults tend to utilize static properties such as the fricative noise spectrum (Nittrouer & Studdert-Kennedy, 1987). This mechanism of perceptual tuning can be directly observed through behavioral experiments, such as when the background noise is a female speaker, the system automatically boosts sensitivity to low-frequency acoustic features of male speech (Mattys et al., 2005). Follow-up studies have found that as language experience accumulates, children complete a stage-by-stage transition in their perceptual strategies between the ages of 6 and 8, gradually developing adult-like processing patterns. The nature of this transition is a dynamic process in which the attentional system optimizes the weighting of acoustic cues through language experience. The longitudinal study by Smith et al. (2005) provides direct evidence for this: they found that 5-7 year olds relied more on dynamic acoustic cues (e.g., resonance peak transitions), whereas 9-11 year olds progressively shifted to static cues (e.g., fricative spectral centers of gravity), and that the speed of the transition was significantly and positively correlated with vocabulary growth (r=0.62, p<0.01), providing strong evidence that the accumulation of language experience drives the attentional system to adaptively reconfigure its processing strategy for acoustic cues, thereby enhancing speech parsing efficiency in the presence of noise (Mattys et al., 2005). This developmental change confirms the dual function of the attentional system: it can adjust the perceptual weights in real time to adapt to the environment, and it can continuously optimize the strategy through experience accumulation.

2.2 Real-time Regulation of Lexical Competition: Suppression of Semantic Interference Driven by Inhibitory Control

After acquiring acoustic cues, language comprehension enters the more challenging phase of lexical selection. At this point, the cognitive system needs to suppress phonetically similar word interference through dynamic regulatory mechanisms. Behavioral experiments have shown that the reaction time when subjects lock onto a target picture in the focused state is significantly shorter than during distracted attention (F(1,30)=12.45, p=0.001), and that this difference in efficiency directly affects conversational fluency (McMurray et al., 2010). However, the cost of modulation failure

is significant: recognition accuracy for the target word " 知道 (zhī dao)" showed significant between-group differences when the near-phonetic word " 支付 (zhī fu)" was present in the background (e.g., the Chinese near-homophones " 知道 (zhī dào, to know)" and " 支付 (zhī fù, to pay)") (Brouwer et al., 2012). McMurray's (2010) eye-tracking experiments found that this inhibition failure was particularly pronounced in the low working-memory capacity group, which showed a decrease in recognition accuracy of up to 27% (Yee & Heller, 2012). This individual difference stems from the interaction between working memory (WM) and inhibitory control (IC). This dynamic regulatory mechanism can be directly observed through eye-tracking experiments: those with higher working memory capacity focus their eyes on grammatical markers such as "although" and "but" when reading complex sentences, whereas those with lower capacity experience frequent retrospection (Gibson, 1998). It has been found that low WM capacity individuals are more disturbed under memory load (Yee & Heller, 2012). Further experiments showed that inhibitory control significantly moderated the anti-interference efficacy of working memory, with low inhibitors taking longer to suppress interfering word activation (Novick et al., 2009). Notably, this interaction effect was dynamic in the temporal dimension: eye-movement data showed that the low ability group took longer to complete suppression, whereas the high ability group could achieve it quickly (Novick et al., 2009).

Collectively, these findings on lexical competition dynamics align with Treisman's (1964) attenuation model, wherein unattended words undergo partial semantic processing. Crucially, this mechanism differs fundamentally from Broadbent's (1958) early filtering model:

(1) Broadbent's model operates at the pre-lexical stage, acting as a "gatekeeper" to filter irrelevant acoustic streams (e.g., suppressing background noise based on pitch differences, as described in Section 2.1).

(2) Treisman's model explains post-lexical interference, where semantically related competitors (e.g., "dog" when targeting "cat") partially activate despite attenuation, necessitating the inhibitory control mechanisms observed above.

This hierarchical division clarifies the complementary roles of both models: Broadbent's filtering governs early acoustic selection, while Treisman's attenuation accounts for mid-stage lexical competition requiring suppression. Consistent with this framework, Treisman's (1964) model of attenuation suggests that unattended words (e.g., near-syntactic words in the background) are processed by attenuation but may still partially activate semantic representations (Treisman, 1964). For example, in a lexical selection task, when the target word is "cat", the interfering word "dog" (in the same animal category) in the background triggers stronger post-decay activation due to conceptual relatedness, requiring the invocation of stronger inhibitory control to suppress the interference (Roelofs, 1992). This model provides a classical theoretical framework for real-time arbitration in lexical competition.

2.3 Strategic Allocation of Syntactic Resources: Working Memory-Driven Parsing of Complex Structures

When dealing with complex sentences, the final stage of speech comprehension requires the invocation of syntactic parsing resources, a process that poses a serious challenge to working memory capacity. When subjects were asked to memorize strings of numbers while comprehending complex sentences such as "The boy who was chased by the dog cried", the error rate increased significantly (Rönnberg et al., 2013). When parsing core syntactic components (e.g., subject-predicate constructions) takes up too many working memory resources, the efficiency of processing grammatical structures in background conversations decreases significantly (Caplan & Waters, 1999), and this competition for resources is further amplified in noisy environments. According to the capacity constraint theory (Just & Carpenter,

1992), when faced with simple subject-verb structures such as "a dog chasing a boy", the brain prioritizes allocating working memory resources to core grammatical components. However, when encountering complex sentence structures with contrasting relationships, such as "Although the boy was chased by the dog, he did not cry", additional executive control resources must be mobilized.

According to the capacity limitation theory (Just & Carpenter, 1992), processing simple structures (e.g., "the dog chased the boy") primarily taxes working memory for core grammatical constituents. In contrast, complex structures involving contrasts (e.g., "Although the boy was chased by the dog, he did not cry") demand additional executive control resources. Eye-tracking experiments have vividly demonstrated this mechanism: people with larger working memories focus on key grammatical markers (e.g., "although" and "but") when reading complex sentences, whereas people with smaller working memories show frequent gaze-back behavior (Gibson, 1998). This difference explains a typical phenomenon in everyday communication: in noisy classrooms, students often hear words clearly but tend to misinterpret "although..." and "but...", and other logical associations of transitive relations.

3 Attentional Control Plasticity from an Individual Differences Perspective

Attentional Control Plasticity from an Individual Differences Perspective reveals that the ability to dynamically regulate the attentional system is not fixed, but shows significant plasticity through second language acquisition, lifespan development, and clinical interventions. Second language learners need to inhibit native language interference through active regulation (Flege, 1995), older bilinguals mitigate auditory decline through strategic compensation (Bak et al., 2014), and children with SLI reshape attentional guidance mechanisms through rhythmic training (Gillam et al., 2008). Together, these cases suggest that plasticity of attentional control is realized through three pathways:

(1) Gradient development of inhibition (from bilingual learners to highly proficient speakers);

(2) Cross-lifespan migration of vicarious strategies (executive function deficits in childhood → bilingual dominance in old age);

(3) Neurobehavioral synergies of cross-boundary training (musical rhythm training → diacritic phonemic perception).

These individual differences in plasticity directly manifest across the hierarchical levels of attentional control: Variations in early acoustic tuning (e.g., diacritic phonemic perception thresholds) influence the efficiency of mid-stage lexical competition (e.g., speed of suppression of near-syllabic words), which in turn shapes late-stage syntactic resource allocation patterns (e.g., working memory capacity). The characteristics of attentional control plasticity and its pedagogical implications for each of the three typical groups will be analyzed below.

3.1 Cognitive Challenges for Second Language Learners

Native speakers process language with their attention automatically targeting key phonological cues through automated mechanisms of attention allocation; whereas second language learners need to extract target speech through active attentional modulation, a process that requires continuous management of competition between the native (L1) and target (L2) languages. It has been found that second language learners have significantly longer reaction times than native speakers for recognizing target words in background noise (Bradlow & Alexander, 2007) and are more susceptible to proxemic word interference. This difference stems from the need for the attentional system to continually inhibit

automatic activation of the native phonological system. Notably, inhibition was significantly and positively correlated with second-language proficiency (r=0.58, p<0.05): high proficiency learners block native language interference more efficiently (Flege, 1995). In addition, inhibitory control training has a transfer effect: Costa et al. (2009) increased monolinguals' word recognition accuracy in noisy environments by 19% (SD=4.2) with 6 weeks of training, an effect that persisted after the training ended (Costa et al., 2009). These findings provide empirical evidence for second language teaching, such as designing Chinese tone contrast training or English burst tone discrimination tasks in noise (Zhao et al., 2017).

According to Flege's (1995) theory of language transfer, automatic activation of the native phonological system interferes with second language processing, and this interference needs to be suppressed by attentional control. For example, when native Chinese speakers acquire English /r/-/l/ pronunciation, their attentional system needs to actively inhibit the native phonemic representation in order to establish a new bilingual phonemic perception, due to the lack of corresponding phonemes in Chinese (Flege, 1995). This theory explains why second language learners need to strengthen the attentional weighting of target phonemes through targeted training.

3.2 Vicarious Mechanisms in the Lifespan

Children aged 5-7 years are unable to effectively utilize intonational cues to filter noise due to underdeveloped executive functions, resulting in significantly higher rates of instruction-following errors (Leibold & Buss, 2013). In contrast, bilingual elders over 65 years of age have significantly higher language comprehension accuracy in the presence of noise than their monolingual peers through compensatory enhancement of attentional control strategies (Bak et al., 2014). This compensatory effect suggests that cognitive plasticity can effectively mitigate the effects of age-related auditory decline on language comprehension. The meta-analytic systematic review of 23 cross-age studies by Schneider et al. (2010) found that older bilinguals maintained higher levels of attentional control through compensatory mechanisms. For example, in a multimodal task requiring simultaneous processing of speech and visual information, older bilinguals showed less attentional distraction than monolinguals (F(2,90)=8.76, p=0.003), and their performance was close to that of younger adults. This vicarious effect may be achieved through the optimization of cognitive control strategies, which at the behavioral level is mainly manifested as an increase in the efficiency of attentional allocation (Schneider et al., 2010).

3.3 Cross-boundary Migration of Clinical Interventions

A core deficit in children with specific language impairment (SLI) is the inability to translate rhyming features into attention-guiding signals. With 8 weeks of rhythmic attention training, children with SLI showed a significant improvement in sentence repetition accuracy in the presence of classroom noise (Stevens et al., 2008). Notably, this improvement manifested itself as a training-specific effect: children who received rhythmic training showed elevated performance on a speech perception task, but no significant change in morphological syntactic processing ability (Corriveau et al., 2007). More strikingly, musical training resulted in a significant increase in children's syllable boundary recognition accuracy by enhancing temporal discrimination, an ability that can be transferred to second language phonological acquisition (Slater et al., 2014).

The effectiveness of rhythmic training is further supported by a clinical study by Gillam et al. (2008). They designed an intervention program for children with SLI that included rhythmic imitation and intonation perception,

and showed that participating children were 72% correct on a sentence comprehension task in a noisy environment, significantly higher than the control group's 51% (t(38)=3.14, p=0.003), and that this improvement persisted after the training ended. This study confirms that reinforcing the attention-guiding effects of rhyming features through targeted training is effective in enhancing the actual communication skills of children with language disorders (Gillam et al., 2008).

4 From Theory to Practice: The Design of Attention-Controlled Training Programs for Language Teaching

Language teaching can actively draw on selective attention theory to design more ecologically valid training programs. For example, the controlled noise training method proposed by Eysenck (1982) requires learners to extract target speech sounds in the midst of distractions by simulating real communication scenarios (e.g., background sounds of a coffee shop) (Eysenck, 1982). Specific task designs may include:

(1) Speech recognition task: students are required to complete the task of recognizing the /v/ and /b/ minimal opposition pairs (e.g., "vest/best") in a noisy environment containing a French background conversation and immediately repeat the target words;

(2) Multimodal reinforcement training: combining visual cues (e.g., speaker mouthing) and haptic feedback (e.g., tapping the wrist to cue the site of articulation) to enhance cross-channel synergy in attentional guidance (Arnold & Hill, 2001).

Experiments have demonstrated that second language learners trained in this way recognize words in noisy environments 25% faster than a traditional clear audio training group (Rost, 2011), and that this effect persists beyond the end of training (F(1,40)=7.32, p=0.01). This training regimen is highly compatible with the "Authentic Corpus Listening" module of the course and can be implemented through the following steps:

(1) Background noise selection: select recordings of real-life scenarios related to the target language and culture (e.g., French café conversations), and ensure that the noise intensity is controlled at 60-65 dB SPL (in line with everyday communication scenarios);

(2) Key Phonological Contrast Design: select confusable phoneme pairs (e.g., /v/ vs. /b/) according to learners' native language background (e.g., Chinese native speakers) and design high-frequency lexical tasks that include these phonemes;

(3) Multimodal cue integration: presenting a video of the speaker's mouth shape synchronized with the listening task, or providing tactile cues to the articulatory site (e.g., the experimenter gently taps the learner's lower lip to indicate the labiodental contact for /v/, or uses a touch-sensitive device to stimulate the lip/teeth area).

Notably, the training effect can be maximized by progressive difficulty tuning: initially using clear speech + light background noise (50 dB SPL), gradually transitioning to real corpus + strong background noise (70 dB SPL), and eventually introducing distractor tasks (e.g., recording key information while listening) in noisy environments. This stepwise design is in line with the hierarchical processing model of attentional control (early acoustic tuning → mid-term lexical tuning → late syntactic assignment), and can be directly applied to the "Teaching Second Language Listening" unit of the "Theory and Practice of Language Teaching" module of the undergraduate program.

5 Conclusion

Language comprehension under noise relies on the dynamic regulation of attentional control: optimizing speech perception through acoustic cues in the early stage, fending off lexical interference with the help of inhibitory control in the middle stage, and safeguarding syntactic integration by relying on working memory in the late stage. Its plasticity feature is the core finding—bilingual experience effectively mitigates auditory decline in old age through inhibitory control, and targeted training (e.g., rhyme intervention) significantly improves sentence comprehension in children with language disorders; and a controlled noise training method that simulates real-life scenarios (e.g., background sounds in a café) has been shown to increase second language learners' word recognition speed in noisy environments by 25% as evidenced by Rost's (2011) reaction time measurements, and the training effect is sustainable. Future research could further explore the cross-domain transfer mechanisms, such as the potential effects of musical rhythm training on L2 conjugation prediction, in order to expand the practical boundaries of cognitive plasticity in language teaching and learning.

References

[1] Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979.

[2] Mattys, S. L. (2012). Speech perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(6), 629–642.

[3] Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*(2), 145–160.

[4] Escudero, P., Birdsong, D., Rota, G., et al. (2016). Age effects in L2 phoneme perception. *Journal of Phonetics*, *54*, 68–79.

[5] Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

[6] McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2010). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *114*(2), 162–173.

[7] Rönnberg, J., Lunner, T., Zekveld, A., et al. (2013). The Ease of Language Understanding model. *Frontiers in Systems Neuroscience*, *7*, 1–15.

[8] Mattys, S. L., Brooks, J., & Cooke, M. (2005). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *51*(2), 141–176.

[9] Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research*, *30*(3), 319–329.

[10] Smith, L. B., Jones, S. S., Landau, B., et al. (2005). Object name learning provides on-the-job training for attention. *Psychological Science*, *16*(1), 13–19.

[11] Brouwer, S., Van Engen, K. J., Calandruccio, L., et al. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners. *The Journal of the Acoustical Society of America*, *132*(4), 2221–2232.

[12] Yee, E., & Heller, D. (2012). Word recognition in the bilingual brain: Evidence from eye-tracking. *Bilingualism: Language and Cognition*, *15*(3), 462–471.

[13] Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

[14] Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(3), 263–281.

[15] Treisman, A. M. (1964). Selective attention in man. *British Medical Bulletin*, *20*(1), 12–16.

[16] Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1–3), 107–142.

[17] Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–94.

[18] Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149.

[19] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

[20] Bak, T. H., Nitzan-Sobel, J., Allerhand, M., et al. (2014). Does bilingualism influence cognitive aging? *Annals of Neurology*, *75*(6), 959–963.

[21] Gillam, R. B., Loeb, D. F., Hoffman, L. M., et al. (2008). The efficacy of Fast ForWord language intervention in school-age children with language impairment: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research*, *51*(1), 97–119.

[22] Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339–2349.

[23] Costa, A., Hernández, M., & Sebastián-Gallés, N. (2009). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*(1), 59–86.

[24] Zhao, T. C., Kulkarni, V., & Giraud, A. L. (2017). Neural mechanisms for coping with acoustically variable speech. *Frontiers in Neuroscience*, *11*, 479.

[25] Leibold, L. J., & Buss, E. (2013). Children's identification of consonants in a speech-shaped noise or a two-talker masker. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1144–1155.

[26] Schneider, B. A., Daneman, M., & Murphy, D. R. (2010). Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? *Psychology and Aging*, *25*(3), 765–777.

[27] Stevens, C., Fanning, J., Coch, D., et al. (2008). Neural mechanisms of selective auditory attention are enhanced by computerized training. *Neuron*, *59*(5), 864–875.

[28] Corriveau, K., Pasquini, E., & Goswami, U. (2007). Basic auditory processing skills and specific language impairment: A new look at an old hypothesis. *Journal of Speech, Language, and Hearing Research*, *50*(3), 647–666.

[29] Slater, J., Tierney, A., & Kraus, N. (2014). Music training improves beat-keeping. *PLOS ONE*, *9*(11), e112466.

[30] Eysenck, M. W. (1982). *Attention and arousal: Cognition and performance*. Berlin: Springer.

[31] Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*(2), 339–355.

[32] Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Harlow: Pearson.