

The Effect of Peripheral Compression on Syllable Perception Measured with a Hearing Impairment Simulator

Toshie Matsui, Toshio Irino, Misaki Nagae, Hideki Kawahara
and Roy D. Patterson

Abstract Hearing impaired (HI) people often have difficulty understanding speech in multi-speaker or noisy environments. With HI listeners, however, it is often difficult to specify which stage, or stages, of auditory processing are responsible for the deficit. There might also be cognitive problems associated with age. In this paper, a HI simulator, based on the dynamic, compressive gammachirp (dcGC) filterbank, was used to measure the effect of a loss of compression on syllable recognition. The HI simulator can counteract the cochlear compression in normal hearing (NH) listeners and, thereby, isolate the deficit associated with a loss of compression in speech perception. Listeners were required to identify the second syllable in a three-syllable “nonsense word”, and between trials, the relative level of the second syllable was varied, or the level of the entire sequence was varied. The difference between the Speech Reception Threshold (SRT) in these two conditions reveals the effect of compression on speech perception. The HI simulator adjusted a NH listener’s compression to that of the “average 80-year old” with either normal compression or complete loss of compression. A reference condition was included where the HI simulator applied a simple 30-dB reduction in stimulus level. The

T. Matsui (✉) · T. Irino · M. Nagae · H. Kawahara
Faculty of Systems Engineering, Wakayama University, Sakaedani 930, Wakayama-city 640-8510,
Japan

e-mail: tmatsui@sys.wakayama-u.ac.jp

T. Irino

e-mail: irino@sys.wakayama-u.ac.jp

M. Nagae

e-mail: s155029@center.wakayama-u.ac.jp

H. Kawahara

e-mail: kawahara@sys.wakayama-u.ac.jp

R. D. Patterson

Department of Physiology, Development and Neuroscience, Centre for the Neural Basis of
Hearing, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK

e-mail: rdp1@cam.ac.uk

© The Author(s) 2016

P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, Advances in Experimental Medicine and Biology 894,
DOI 10.1007/978-3-319-25474-6_32

results show that the loss of compression has its largest effect on recognition when the second syllable is attenuated relative to the first and third syllables. This is probably because the internal level of the second syllable is attenuated proportionately more when there is a loss of compression.

Keywords Hearing impairment simulator · Cochlear compression · Dynamic compressive gammachirp filterbank · Audiogram

1 Introduction

Age related hearing loss (presbycusis) makes it difficult to understand speech in noisy environments and multi-speaker environments (Moore 1995; Humes and Dubno 2010). There are several factors involved in presbycusis: loss of frequency selectivity, recruitment, and loss of temporal fine structure (Moore 2007). The isolation and measurement of these different aspects of hearing impairment (HI) in elderly listeners is not straightforward, partly because they often have multiple auditory problems, and partly because they may also have more central cognitive problems including depression. Together the problems make it difficult to obtain sufficient behavioral data to isolate different aspects of HI (Jerger et al. 1989; Lopez et al. 2003). We have developed a computational model of HI and an HI simulator that allow us to perform behavioural experiments on cochlear compression—experiments in which young normal hearing (NH) listeners act as “patients” with isolated, sensory-neural hearing losses of varying degree.

The HI simulator is essentially a high-fidelity synthesizer based on the dynamic, compressive gammachirp filter bank (dcGC-FB) (Irino and Patterson 2006). It analyses naturally recorded speech sounds with a model of the auditory periphery that includes fast-acting compression and then resynthesizes the sounds in a form that counteracts the compression in a NH listener. They hear speech with little distortion or background noise (Irino et al. 2013) but without their normal compression. This makes it possible to investigate recruitment phenomenon that are closely related to loss of cochlear compression (Bacon et al. 2004) using NH listeners who have no other auditory problems. Specifically, we describe an experiment designed to examine how the loss of cochlear compression affects the recognition of relatively soft syllables occurring in the presence of louder, flanking syllables of varying levels. The aim is to reveal the role of cochlear compression in multi-speaker environments or those with disruptive background noise.

2 Method

Listeners were presented three-syllable “nonsense words” and required to identify the second syllable. Between trails, the level of the second syllable within the word was varied (“dip” condition) or the level of the entire word was varied (“constant”

condition). The difference between second-syllable recognition in the two conditions reveals the effect of compression. The HI simulator is used to adjust syllable level to simulate the hearing of 80 year old listeners with “normal” presbycusis (80 year 100%) or a complete loss of compression (80 year 0%).

2.1 Participants

Ten listeners participated in the experiment (four males; average age: 23.6 years; SD: 4.8 years). None of the listeners reported any history of hearing impairment. The experiment was approved by the ethics committee of Wakayama University and all listeners provided informed consent. Participants were paid for their participation, except for the first and third authors who also participated in the experiment.

2.2 Stimuli and Procedure

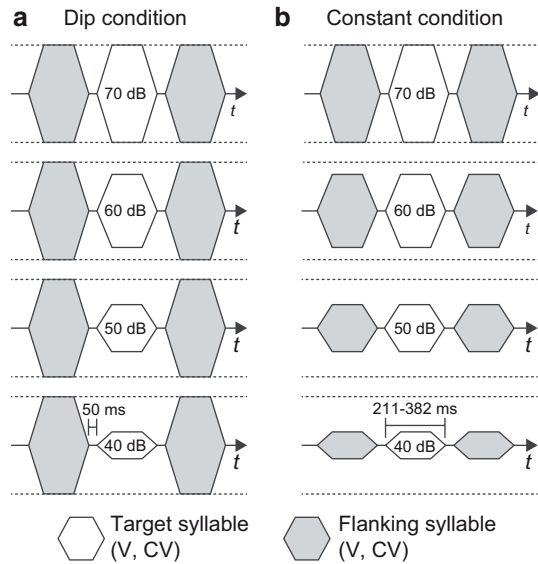
The stimuli were generated in two stages: First, three-syllable nonsense words were composed from the recordings of a male speaker, identified as “MIS” (FW03) in a speech sound database (Amano et al. 2006), and adjusted in level to produce the dip and constant conditions of the experiment. Second, the HI simulator was used to modify the stimuli to simulate the hearing of two 80 year old listeners, one with average hearing for an 80 year old and one with a complete loss of compression. There was also a normal-hearing control condition in which the level was reduced a fixed 30 dB.

In the first stage, nonsense words were composed by choosing the second syllable at random from the 50 Japanese syllables presented without parentheses in Table 1. This is the list of syllables (57-S) recommended by the Japanese Audiological Society (2003) for use in clinical studies to facilitate comparison of behavioural data across studies. The first and third syllables in the words were then selected at random from the full set of 62 syllables in Table 1. The set was enlarged to reduce the chance of listeners recognizing the restriction on the set used for the target

Table 1 The 62 Japanese syllables used in the experiment using the international phonetic alphabet. The syllables in parentheses are also Japanese syllables but not included in list 57-S. All 62 were used in the draw for the first and third syllables. The draw for the second syllable was restricted to the 50 syllables not enclosed in parentheses

a	ka	sa	ta	na	ha	ma	ja	ra	wa	ga	(dza)	da	ba
i	kɨ	ʃi	tʃi	ɲi	çi	mi		ɾi		(gi)	dʒi		(bi)
u	ku	su	tsu	(nu)	ɸu	mu	ju	ru		(gu)	dzu		(bu)
e	ke	se	te	ne	he	me		(re)		(ge)	(dze)	de	(be)
o	ko	so	to	no	ho	mo	jo	ro		go	(dzo)	do	(bo)

Fig. 1 Schematic of the experimental paradigm. The levels of the nonsense words were adjusted to produce stimuli for **a** the Dip condition and **b** the Constant condition. The 70-dB stimuli were the same in the two conditions



syllable. Syllable duration ranged from 211 to 382 ms. There was a 50-ms silence between adjacent syllables within a word.

The nonsense words were then adjusted to produce the two level conditions illustrated in Fig. 1: In the “dip” condition (left column), the level of the second syllable was 70, 60, 50 or 40 dB, while the first and third syllables were fixed at 70 dB. In the “constant” condition (right column), all three syllables were set to 70, 60, 50 or 40 dB.

In the second stage, the stimuli were processed to simulate:

1. the average hearing level of an 80 year old person (audiogram ISO7029, ISO/TC43 2000) simulated with 100% compression (80 year 100%),
2. the average hearing level of an 80-year-old simulated with 0% compression (80 year 0%),
3. normal adult hearing with a 30-dB reduction across the entire frequency range (NH – 30 dB).

To realize the 80 year 100% condition, the output level of each channel was decreased until the audiogram of the dcGC-FB (normal hearing) matched the average audiogram of an 80 year old. To realize the 80 year 0% condition, an audiogram was derived with the dcGC-FB having no compression whatsoever, and then the output level of each channel was decreased until the audiogram matched the audiogram of an average 80 year old. In the compression-cancellation process, loss of hearing level was limited so that the audiogram did not fall below the hearing level of an average 80 year old. The compression was applied using the signal processing system described in Nagae et al. (2014).

Each participant performed all conditions—a total of 1050 trials of nonsense words: 50 words \times 7 dip/constant conditions \times 3 hearing impairment simulations. The participants were required to identify the second syllable using a GUI on a computer screen. The stimuli were manipulated, and the results collected using MATLAB. The sounds were presented diotically via a DA converter (Fostex, HP-A8) over headphones (Sennheiser, HD-580). The experiment was carried out in a sound-attenuated room with a background level of about 26 dB L_{Aeq} .

3 Results

For each participant, the four percent-correct values for each level condition (dip or constant) were fitted with a cumulative Gaussian psychometric function to provide a Speech Reception Threshold (SRT) value for that condition, separately in each of the three simulated hearing-impairment conditions. The cumulative Gaussian was fitted by the bootstrap method (Wichmann and Hill 2011a; Wichmann and Hill 2011b) and SRT was taken to be the sound pressure level associated with 50% correct recognition of the second syllable. The SRT in the constant condition with the NH - 30 dB hearing impairment was taken as standard performance on the task. The difference between this standard SRT and the SRTs of the other conditions are the experimental results (Δ SRTs) for each participant. Figure 2a shows the mean and standard deviation over participants of Δ SRT for each condition.

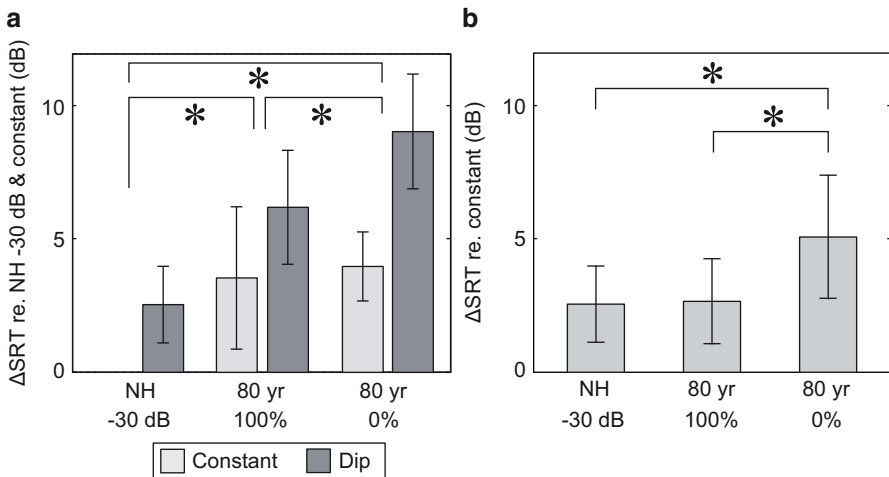


Fig. 2 **a** The difference in SRT (Δ SRT) between the constant condition of the NH - 30 dB simulation and the SRTs of the other conditions. **b** The SRT difference between the dip and constant conditions for each hearing impairment condition. Asterisks (*) show significant differences ($\alpha=0.05$) in a multiple comparison, Tukey-Kramer HSD test. Error bars show the standard deviation

A two-way ANOVA was performed for the dip/constant conditions and the hearing impairment conditions combined, and it showed that the manipulations had a significant effect on recognition performance (Δ SRT) ($F(5, 54)=23.82, p<0.0001$). Significant, separate main effects were also observed for the dip/constant condition and the hearing impairment condition ($F(1, 54)=41.99, p<0.0001$; $F(2, 54)=36.11, p<0.0001$, respectively). The interaction between the two factors was not significant ($\alpha=0.05$). Multiple comparisons with the Tukey-Kramer HSD test were performed. There were significant differences between all of the hearing impairment conditions ($\alpha=0.05$).

Δ SRT provides a measure of the difficulty of the task, and the statistical analysis shows that the different hearing impairment simulations vary in their difficulty. In addition, Δ SRT for the dip conditions is larger than Δ SRT for constant conditions for all of the hearing impairment simulations. This difference in difficulty could be the result of forward masking since the gap between offset of the first syllable and the onset of the target syllable is less than 100 ms, and the third syllable limits any post processing of the second syllable. The dip condition would be expected to produce more masking than the constant condition, especially when the target syllable is sandwiched between syllables whose levels are 20 or 30 dB greater. The constant condition only reveals differences between the different hearing impairment simulations. This does, however, include differences in audibility.

Figure 2b shows the difference between SRT in the dip and constant conditions; that is, Δ SRT relative to the constant condition. It provides a measure of the amount of any forward masking. A one-way ANOVA was performed to confirm the effect of hearing impairment simulation ($F(2, 29)=6.14, p=0.0063$). A multiple comparisons test (Tukey-Kramer HSD) showed that Δ SRT in the 80 year 0% condition was significantly greater than in the 80 year 100% condition, or the NH -30 dB condition ($\alpha=0.05$).

4 Discussion

It is assumed that forward masking occurs during auditory neural processing, and that it operates on peripheral auditory output (i.e. neural firing in the auditory nerve). Since the compression component of the peripheral processing is normal in both the 80 year 100% condition and the NH -30 dB condition, the level difference between syllable 2 and syllables 1 and 3 should be comparable in the peripheral output of any specific condition. In this case, Δ SRT provides a measure of the more central masking that occurs during neural processing. When the compression is cancelled, the level difference in the peripheral output for syllable 2 is greater in the dip conditions. If the central masking is constant, independent of the auditory representation, the reduction of the output level for syllable 2 should be reflected in the amount of masking, i.e. Δ SRT. This is like the description of the growth of masking by Moore (2012), where the growth of masking in a compressive system is observed to be slower than in a non-compressive system. Moore was describing experiments

where both the signal and the masker were isolated sinusoids. The results of the current study show that the effect generalizes to speech perception where both the target and the interference are broadband, time-vary sounds.

There are, of course, other factors that might have affected recognition performance: The fact that the same speaker provided all of the experimental stimuli might have reduced the distinctiveness of the target syllable somewhat, especially since syllable 2 was presented at a lower stimulus level in the dip conditions. The fact that listeners had to type in their responses might also have decreased recognition performance a little compared to everyday listening. It is also the case that our randomly generated nonsense words might accidentally produce a syllable sequence that is similar to a real word with the result that the participant answers with reference to their mental dictionary and makes an error. But, all of these potential difficulties apply equally to all of the HI conditions, so it is unlikely that they would affect the results.

This suggests that the results truly reflect the role of the loss of compression when listening to speech in everyday environments for people with presbycusis or other forms of hearing impairment. It shows that this new, model-based approach to the simulation of hearing impairment can be used to isolate the compression component of auditory signal processing. It may also assist with isolation and analysis of other aspects of hearing impairment.

5 Summary

This paper reports an experiment performed to reveal the effects of a loss of compression on the recognition of syllables presented with fluctuating level. The procedure involves using a filterbank with dynamic compression to simulate hearing impairment in normal hearing listeners. The results indicate that, when the compression of normal hearing has been cancelled, it is difficult to hear syllables at low sound levels when they are sandwiched between syllables with higher levels. The hearing losses observed in patients will commonly involve multiple auditory problems and they will interact in complicated ways that make it difficult to distinguish the effects of any one component of auditory processing. The HI simulator shows how a single factor can be isolated using an auditory model, provided the model is sufficiently detailed.

Acknowledgments This work was supported in part by JSPS Grants-in-Aid for Scientific Research (Nos 24300073, 24343070, 25280063 and 25462652).

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Amano S, Sakamoto S, Kondo T, Suzuki Y (2006) NTT-Tohoku University familiarity controlled word lists (FW03). Speech resources consortium. National Institute of Informatics, Japan
- Bacon SP, Fay R, Popper AN (eds) (2004) *Compression: from cochlea to cochlear implants*. Springer, New York
- Humes LE, Dubno JR (2010) Factors affecting speech understanding in older adults. In: Gordon-Salant S, Frisina RD, Popper AN, Fay RR (eds) *The aging auditory system*. Springer, New York, pp 211–257
- Irino T, Patterson RD (2006) A dynamic compressive gammachirp auditory filterbank. *IEEE Trans Audio Speech Lang Process* 14(6):2222–2232
- Irino T, Fukawatase T, Sakaguchi M, Nisimura R, Kawahara H, Patterson RD (2013) Accurate estimation of compression in simultaneous masking enables the simulation of hearing impairment for normal hearing listeners. In: Moore BCJ, Patterson RD, Winter IM, Carlyon RP, Gockel HE (eds) *Basic aspects of hearing, physiology and perception*. Springer, New York, pp 73–80
- ISO/TC43 (2000) *Acoustics—statistical distribution of hearing thresholds as a function of age*. International Organization for Standardization, Geneva
- Japan Audiological Society (2003) The 57-S syllable list. *Audiol Jpn* 46:622–637
- Jerger J, Jerger S, Oliver T, Pirozzolo F (1989) Speech understanding in the elderly. *Ear Hear* 10(2):79–89
- Lopez OL, Jagust WJ, DeKosky ST, Becker JT, Fitzpatrick A, Dulberg C, Breitner J, Lyketsos C, Jones B, Kawas C, Carlson M, Kuller LH (2003) Prevalence and classification of mild cognitive impairment in the cardiovascular health study cognition study: part 1. *Arch Neurol* 60(10):1385–1389
- Moore BCJ (1995) *Perceptual consequences of cochlear damage*. Oxford University Press, New York
- Moore BCJ (2007) *Cochlear hearing loss: physiological, psychological and technical issues*. Wiley, Chichester
- Moore BCJ (2012) *An introduction to the psychology of hearing*, 6th edn. Emerald Group Publishing Limited, Bingley
- Nagae M, Irino T, Nishimura R, Kawahara H, Patterson RD (2014). Hearing impairment simulator based on compressive gammachirp filter. *Proceedings of APSIPA ASC*
- Wichmann FA, Hill NJ (2011a) The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept Psychophys* 63(8):1293–1313
- Wichmann FA, Hill NJ (2011b) The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept Psychophys* 63(8):1314–1329