Predicting effects of hearing-instrument signal processing on consonant perception

Zaar, Johannes; Schmitt, Nicola; Derleth, Ralph-Peter; DiNino, Mishaela; Arenberg, Julie G.; Dau, Torsten

Published in:
Journal of the Acoustical Society of America

Link to article, DOI:
10.1121/1.5011737

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Zaar, J., Schmitt, N., Derleth, R.-P., DiNino, M., Arenberg, J. G., & Dau, T. (2017). Predicting effects of hearing-instrument signal processing on consonant perception. Journal of the Acoustical Society of America, 142(5), 3216-26. https://doi.org/10.1121/1.5011737

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Predicting effects of hearing-instrument signal processing on consonant perception
Johannes Zaar, Nicola Schmitt, Ralph-Peter Derleth, Mishaela DiNino, Julie G. Arenberg, and Torsten Dau

Citation: The Journal of the Acoustical Society of America 142, 3216 (2017);
View online: https://doi.org/10.1121/1.5011737
View Table of Contents: http://asa.scitation.org/toc/jas/142/5
Published by the Acoustical Society of America

Articles you may be interested in
Reliability of individual differences in degraded speech perception
The Journal of the Acoustical Society of America 142, EL461 (2017); 10.1121/1.5010148

Cochlear implant simulator with independent representation of the full spiral ganglion
The Journal of the Acoustical Society of America 142, EL484 (2017); 10.1121/1.5009602

Predicting consonant recognition and confusions in normal-hearing listeners
The Journal of the Acoustical Society of America 141, 1051 (2017); 10.1121/1.4976054

Overall judgment of loudness of time-varying sounds
The Journal of the Acoustical Society of America 142, 1841 (2017); 10.1121/1.5003797

Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech
The Journal of the Acoustical Society of America 142, 2823 (2017); 10.1121/1.5008372

Pitch matching in bimodal cochlear implant patients: Effects of frequency, spectral envelope, and level
The Journal of the Acoustical Society of America 142, 2854 (2017); 10.1121/1.5009443
Predicting effects of hearing-instrument signal processing on consonant perception

Johannes Zaara)

Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

Nicola Schmitt and Ralph-Peter Derleth

Sonova AG, Stäfa, CH-8712, Switzerland

Mishaela DiNino and Julie G. Arenberg

Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd Street, Box 354875, Seattle, Washington 98105, USA

Torsten Dau

Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

(Received 27 April 2017; revised 1 September 2017; accepted 31 October 2017; published online 22 November 2017)

This study investigated the influence of hearing-aid (HA) and cochlear-implant (CI) processing on consonant perception in normal-hearing (NH) listeners. Measured data were compared to predictions obtained with a speech perception model [Zaar and Dau (2017). J. Acoust. Soc. Am. 141, 1051–1064] that combines an auditory processing front end with a correlation-based template-matching back end. In terms of HA processing, effects of strong nonlinear frequency compression and impulse-noise suppression were measured in 10 NH listeners using consonant-vowel stimuli. Regarding CI processing, the consonant perception data from DiNino et al. [(2016). J. Acoust. Soc. Am. 140, 4404–4418] were considered, which were obtained with noise-vocoded vowel-consonant-vowel stimuli in 12 NH listeners. The inputs to the model were the same stimuli as were used in the corresponding experiments. The model predictions obtained for the two data sets showed a large agreement with the perceptual data both in terms of consonant recognition and confusions, demonstrating the model’s sensitivity to supra-threshold effects of hearing-instrument signal processing on consonant perception. The results could be useful for the evaluation of hearing-instrument processing strategies, particularly when combined with simulations of individual hearing impairment.

© 2017 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.1121/1.5011737

[ICB] Pages: 3216–3226

I. INTRODUCTION

Speech perception is commonly tested by assessing the percentage of correctly identified words or sentences in the presence of some acoustical interference or degradation, such as additive noise and/or reverberation (cf. Hagerman, 1982; Nilsson et al., 1994; Wagener et al., 2003; Nielsen and Dau, 2009, 2011). While such speech tests provide some useful “macroscopic” information regarding the effects of different acoustic conditions on intelligibility, the typically used speech reception threshold measure (SRT), representing the signal-to-noise ratio (SNR) at which 50% intelligibility is obtained, is rather coarse as it reflects responses averaged across many speech tokens. Furthermore, the listeners can “restore” missing acoustic information using semantic predictability and lexical information (e.g., Miller and Licklider, 1950; Warren, 1970; Bashford et al., 1992; Kashino, 2006), such that linguistic processing ability may strongly influence the listeners’ performance. Moreover, the frequency importance function for the intelligibility of sentences is strongly dominated by the low-frequency speech content (Pavlovic, 1987), such that macroscopic speech intelligibility tests are not very sensitive to effects in the mid and high frequency ranges (e.g., due to high-frequency masking noise, filtering, or nonlinear speech processing). Therefore, such tests may not be well suited for investigating effects of hearing impairment (typically most pronounced at high frequencies) and hearing-instrument signal processing on speech perception.

Instead, it can be insightful to examine the perception of individual phonemes, sometimes referred to as a “microscopic” approach of studying speech perception. Various studies have focused on the perception of consonants embedded in nonsense syllables in normal-hearing (NH) listeners (e.g., Miller and Nicely, 1955; Wang and Bilger, 1973; Phatak et al., 2007; Phatak et al., 2008; Zaar and Dau, 2015), e.g., in the form of consonant-vowel (CV) combinations (/ba/, /ta/, etc.), typically presented in

a)Electronic mail: jzaar@elektro.dtu.dk
steady-state noise at various SNRs. In such tests, the contribution of high-level cognitive restoration effects is eliminated due to the nonsense nature of the stimuli and the importance of the critical high-frequency speech cues is taken into account as many consonant cues contain high-frequency energy (cf. Li et al., 2010; Li et al., 2012). Furthermore, not only can consonant recognition performance be evaluated, but also the patterns of consonant confusions, indicating the types of errors that occurred.

Several studies have investigated the effects of hearing impairment on consonant perception (e.g., Phatak et al., 2009; Trevino and Allen, 2013). Scheidiger et al. (2017) studied the influence of different amplification schemes on consonant perception in hearing-impaired (HI) listeners and demonstrated that consonant perception tests may be more informative for hearing-aid (HA) fitting than pure-tone audiometry. Schmitt et al. (2016) presented a consonant perception test specifically designed for high-frequency HA fitting, which determines (i) audibility thresholds of high-pass filtered representations of /s/ and /ʃ/ and (ii) recognition thresholds of these consonants in a vowel-consonant-vowel (VCV) context (i.e., /a, s, aʃa/). Testing HI listeners with and without HAs, they demonstrated that the test was sensitive to effects of high-frequency amplification as well as nonlinear frequency compression (NLFC). NLFC (Simpson et al., 2005) attempts to restore high-frequency acoustic information in listeners with pronounced high-frequency hearing loss by compressing the high-frequency signal content and shifting it to lower frequencies, as HAs typically cannot provide sufficient gain at frequencies above 5 kHz (Kimlinger et al., 2015). Glista et al. (2009) showed that NLFC can substantially improve high-frequency consonant recognition scores in listeners with a high-frequency hearing loss. However, NLFC with “too strong” settings can result in a drastic reduction of consonant recognition, as demonstrated by Schmitt et al. (2016). This is consistent with the strongly frequency dependent acoustic cues that lead to different consonant percepts (cf. Li et al., 2010; Li et al., 2012), as frequency-compressed high-frequency consonants may perceptually “morph” into other consonants that exhibit a temporally similar cue in a lower frequency region. For example, /s/ and /ʃ/ are represented by friction noise at very high and slightly lower frequencies, respectively, such that a too strong NLFC leads to /s/ being perceived as /ʃ/. However, such perceptual morphs may disappear after an acclimatization period due to re-learning of the modified consonant cues (cf. Wolfe et al., 2011). Consonant perception depends not only on the spectral characteristics of the signal but also on its temporal properties. Temporal signal modifications due to the highly nonlinear processing schemes typically applied in HAs [e.g., impulse-noise suppression (INS)] may thus also affect consonant perception.

An alternative compensation strategy is represented by cochlear implant (CI) processing, applied in more severe cases of hearing impairment. CIs yield great improvements in terms of speech intelligibility by transmitting individual frequency bands of a signal directly to different places in the cochlea using an implanted electrode array. However, CIs are limited with respect to spectral resolution, as the number of electrodes in an implanted array is limited and channel interactions typically occur (White et al., 1984; Stickney et al., 2006). Furthermore, spectral resolution may be further degraded due to poor electrode-neuron interfaces—defined by regions of poor neural survival or a large distance between the CI electrodes and the auditory neurons (for review, see Bierer, 2010). DiNino et al. (2016) investigated the effect of CI processing with poor electrode–neuron interfaces on the perception of consonants and vowels in NH listeners using VCV and consonant-vowel-consonant (CVC) syllables, respectively, that were noise-vocoded to simulate CI processing. A reference CI simulation condition using all available channels was considered along with conditions where low-, middle-, and high-frequency channels were either set to zero (“Zero”), simulating neural dead regions, or re-distributed to neighboring channels (“Split”), simulating poor electrode positioning. While listeners exhibited considerable perceptual differences across the considered frequency regions (but not across the Zero and Split conditions) in the vowel perception test, the consonant perception test showed less variability across frequency regions, as all CI processing conditions induced largely similar effects on consonant perception.

To better understand how various aspects of HA and CI processing affect consonant perception, computational models of speech perception may serve as valuable tools. If such a model can account for the effects of specific HA/CI processing strategies on consonant perception, it may provide useful information about the auditory cues that contribute to the recognition of a specific consonant or its confusion with another consonant. Several approaches for modeling consonant perception in NH listeners (Cooke, 2006; Jürgens and Brand, 2009) and in HI listeners (Holube and Kollmeier, 1996; Jürgens et al., 2014; Jepsen et al., 2014) have been proposed. While the mentioned models were shown to account for consonant recognition scores in masking noise (or in quiet at low signal levels), they did not account well for the consonant confusions, i.e., the predicted errors were different from the listeners’ errors. Hearing-instrument signal processing, on the other hand, may lead to changed rather than masked consonant cues, inducing specific strong consonant confusions (cf. Schmitt et al., 2016; DiNino et al., 2016). A model that can account for such effects therefore needs to be sensitive not only to the presence of a consonant cue, but also to its perceptual similarity with other consonant cues. Zaar and Dau (2017) proposed a consonant perception model that appears to provide such sensitivity. It combines an auditory model (Dau et al., 1996, 1997) that includes adaptive processes and modulation-frequency selective processing with a temporally dynamic correlation-based template-matching back end. The model was evaluated on the extensive data set by Zaar and Dau (2015), obtained in NH listeners with CVs presented in white noise at various SNRs. The model was shown to account well for consonant recognition even at the level of individual speech tokens. Moreover, a good agreement of the model predictions with the perceptual consonant confusions was demonstrated, albeit with some underestimation of the perceptual confusions’ extent.
In the present study, the effects of several HA and CI processing conditions on consonant perception were investigated using the model of Zaar and Dau (2017) to predict the behavioral data. In particular, an experimental investigation of the effects of HA processing (NLFC and INS) on NH listeners’ consonant perception was conducted using speech material from Schmitt et al. (2016). Furthermore, the data from DiNino et al. (2016) were considered, representing effects of simulated CI processing on NH listeners’ consonant perception. Model predictions were obtained for the two data sets based on the respective experimental stimuli. The model performance was evaluated by means of confusion matrix (CM) comparisons, as well as on the basis of correlation analyses of the perceptual and predicted consonant recognition and confusion scores.

II. METHOD

A. Experiment 1: Effects of HA signal processing

1. Stimuli and experimental conditions

The audio material was taken from the speech material recorded by Schmitt et al. (2016) and consisted of the VCVs /aba, aʃa, ada, apa, aka, ata, asa, aʃa, aʃa, ata/1, spoken by a female native German speaker. The speaker was trained to speak all VCVs with similar speed and pitch. Schmitt et al. (2016) used two versions of /asa/ and /aʃa/, respectively, filtered to have different spectral peaks: /ʃ/ exhibited a spectral peak at 4.6 kHz and was spectrally shaped to show spectral peaks at 3 and 5 kHz, resulting in /aʃa/ and /aʃa/. /s/ exhibited a spectral peak at 7.2 kHz and was spectrally shaped to show spectral peaks at 6 and 9 kHz, resulting in /asa/ and /asa/1. For evaluating effects of INS on consonant perception, the stimuli need to start with the consonant. Thus, the initial vowels of the considered 12 VCV tokens /aba, aʃa, ada, apa, aka, ata, asa, aʃa, aʃa, aʃa, ata/ were manually removed to obtain the CVs /ba, fa, da, pa, ka, ta, sa, sa, fa, tʃa, fa, ts/. Five conditions were considered: unaided, default, NLFC, INS, and NLFC&INS. The unaided condition was a natural listening situation without HA processing. For the other four conditions, Phonak Naída V90-RIC HAs were employed, assuming a moderate-to-severe hearing loss with 55 dB hearing level (HL) at frequencies of 1 kHz and below, 65 dB HL at 2 kHz, 75 dB HL at 4 kHz, and 80 dB HL at 8 kHz. For all four HA conditions, NAL-NL2 amplification was selected. The default condition was defined as the default HA settings suggested by the fitting software, which included soft NLFC (using Phonak SoundRecover) that compressed the range between 3.8 and 10 kHz by a factor of 2.4 to the range between 3.8 and 5.7 kHz. In the NLFC condition, the strongest possible setting of Phonak SoundRecover was selected, such that the frequency content in the range between 1.5 and 10 kHz was compressed by a factor of 4 to the range between 1.5 and 2.41 kHz. In the INS condition, the strongest possible setting of the provided impulse-noise suppression (Phonak SoundRelax) was selected. In the NLFC&INS condition, NLFC and INS were combined using the settings described above for the NLFC condition and the INS condition, respectively. For all HA settings, omnidirectional microphone directivity was selected.

One sound file with all CVs was obtained by concatenating the CVs with 500-ms pauses between them. Steady-state speech-shaped noise (SSN) with a long-term average spectrum of female speech was added at an effective SNR of 8 dB (in the speech-containing portions). Ten seconds of noise alone preceded the first CV. The mixture of CVs and noise was played back from a loudspeaker, positioned at a distance of 1.5 m and 0° azimuth relative to a KEMAR dummy head in a sound-attenuating room. The speech level at the position of the dummy head was set to 70 dBA. The signals were recorded at a sampling rate of 48 kHz at the position of the dummy head’s left tympanic membrane either without HA (unaided condition) or with HA using the condition-specific HA setting. The recordings were equalized to compensate for the applied amplification and cut into individual CV stimuli with 350 ms of noise at the beginning and 50 ms of noise at the end, using 50-ms raised-cosine ramps for fade in/out.

2. Listening test

Ten adult NH native German listeners (mean age: 29.5 years; standard deviation: 3.6 years) with audiometric thresholds of 20 dB hearing level (HL) or less between 125 Hz and 8 kHz were tested. The listeners were seated in a sound-insulated booth in front of a computer screen and binaurally presented with the diotic stimuli via Sennheiser HD 650 headphones at 60 dB sound pressure level (SPL). They were asked to select the consonants they heard on a graphical user interface, which displayed the considered response alternatives /b, d, p, k, t, s, ʃ, f, ts/ in the corresponding German spelling (b, d, g, d, k, t, s, sch, f, z). Each of the 60 stimuli (12 CVs in five conditions) was presented eight times to each listener, amounting to a total of 480 stimulus presentations per listener. The order of presentation was randomized across CVs and conditions. After the listener had made a decision, the next stimulus was played after a pause of 500 ms. The experiment duration was about 25 min per listener. No training or feedback was provided to the listeners and the stimulus presentation could not be repeated. As some stimuli sounded rather ambiguous, listeners were instructed to select the response alternative that most closely resembled what they heard. The frequencies of responses obtained for each stimulus were summed across listeners and divided by the overall number of presentations (80: 8 presentations × 10 listeners) to obtain the proportions of responses.

B. Experiment 2: Effects of CI signal processing

DiNino et al. (2016) considered sixteen VCVs, consisting of consonants embedded in an /aCa/ context (/p/, “apa”; /f/, “ata”; /k/, “aka”; /b/, “aba”; /d/, “ada”; /g/, “aga”; /ʃ/, “aʃa”; /l/, “atha”; /s/, “asa”; /ʃ/, “asha”; /ʃ/, “ava”; /z/, “aza”; /dʒ/, “aja”; /n/, “ama”; /n/, “ana”; /l/, “ala”). All VCVs were naturally spoken by a male talker (native speaker of American English). Vocoder processing was applied to the stimuli to simulate CI processing in combination with regions of poor neural survival. The processing was designed to simulate “CI
fidelity 120° processing with the same frequency band allocations as Advanced Bionics devices and realized in MATLAB using a CI simulation software developed by Litvak et al. (2007). Fifteen vocoder bands with logarithmic spacing in the frequency range between 250 Hz and 8.7 kHz and a slope of 30 dB/octave were considered for the simulations. The sub-band envelopes of the VCVs were extracted, lowpass-filtered at 68 Hz, and used to modulate noise bands with the same center frequencies. As a control condition, the VCVs were processed using all vocoder bands (AllChannels). For the other six conditions, the spectral information in three frequency regions (Apical / 421–876 Hz; Middle / 877–1826 Hz; Basal / 1827–3808 Hz) was degraded by either (i) setting the corresponding channels to zero (Zero) or (ii) setting them to zero and adding half of the envelope energy from the zeroed channels to the neighboring lower-frequency channels and the other half to the adjacent higher-frequency channels (Split). The noise bands were summed and the resulting vocoded stimuli were stored at a sampling rate of 17.4 kHz.

Twelve adult NH listeners with a mean age of 25.2 years participated in the study of DiNino et al. (2016). All listeners were native speakers of American English. All 112 VCV stimuli (16 VCVs × 7 conditions) were presented three times in random order to each listener at 60 dBA via a loudspeaker positioned one meter from the subject at 0° azimuth in a sound-insulated booth. Listeners were asked to select the consonant they heard on a computer screen. Two such experimental blocks were run, such that six responses per list were summed across listeners and divided by the overall number of stimulus presentations (72; 6 presentations × 12 listeners) to obtain the proportions of responses.

C. Model simulations

1. Model description

The consonant perception model of Zaar and Dau (2017) was considered for predicting the perceptual data obtained with the HA-processed CVs as well as with the CI-processed VCVs. Figure 1 shows the model, which combines the auditory model front end of Dau et al. (1996, 1997) with a temporally dynamic correlation-based back end. The auditory model consists of (i) a bank of 15 fourth-order gammatone filters with center frequencies logarithmically spaced between 315 Hz and 8 kHz, (ii) an envelope extraction stage (realized by half-wave rectification and lowpass filtering at 1 kHz), (iii) a chain of five adaptation loops (designed to mimic adaptive properties of the auditory periphery), and (iv) a bank of four modulation filters, implemented as a 2-Hz lowpass filter in parallel with three second-order bandpass filters with a Q-factor of 1 and center frequencies of 4, 8, and 16 Hz, respectively. For a given noisy speech signal, the temporal pattern of the noise alone (after the preprocessing stages) is subtracted from the corresponding temporal pattern of the noisy speech. The resulting model representations of the test signal ($R_{test}$) and of a set of templates ($R_{t1}, R_{t2}, \ldots, R_{tn}$) are then aligned in time using a dynamic time warping (DTW) algorithm (Sakoe and Chiba, 1978). Finally, the cross-correlation coefficients between the time-aligned test-signal representation ($R_{test}$) and the time-aligned template representations ($R_{t1}, R_{t2}, \ldots, R_{tn}$) are calculated and, after adding a constant-variance internal noise to limit the model’s resolution, converted to response percentages.

2. Simulation procedure

To predict the data from experiment 1, the recorded HA-processed CVs that were used as experimental stimuli were fed to the model. Portions of the respective dummy-head recordings that contained only noise were considered as “noise alone” signals in the model (depending on the condition of the considered stimulus). The CV recordings obtained in the unaided condition were considered as templates since they had not been passed through a HA but still contained the effects of the noise, the room, and the KEMAR dummy head on the CV speech tokens. Nine templates were generated from each noisy CV recording by using nine randomly selected samples of the noise alone, such that the template-matching procedure could be iterated.

FIG. 1. Scheme of the consonant perception model (reprinted from Zaar and Dau, 2017). For the test signal and a set of templates, the noisy speech and the noise alone were passed separately through the auditory model, consisting of a gammatone filterbank, an envelope extraction stage, a chain of adaptation loops, and a modulation filterbank. The difference between the temporal patterns of the noisy speech and the noise alone was obtained. The resulting representations of the test signal and the templates were time-aligned using a DTW algorithm. Finally, the cross-correlation coefficients between the test signal and each template were calculated and, after addition of a constant-variance internal noise, converted to percent.
nine times. After obtaining the correlation coefficients between each test signal and all templates, the internal noise was added and the model response for each iteration was defined as the template showing the largest correlation with the test signal. As proposed in Zaar and Dau (2017), the model was calibrated by adjusting the variance of the internal noise based on the average consonant recognition scores obtained in the considered conditions. Here, a variance of $\sigma^2_{int,1} = 0.15$ was found to be optimal, which is larger than the variance of 0.05 used in Zaar and Dau (2017). This larger internal noise was necessary to account for the higher difficulty that the listeners experienced due to the HA signal processing conditions considered here (as compared to the additive white noise conditions considered in Zaar and Dau, 2017). However, the internal noise was held constant across the considered conditions. For each test signal, the numbers of occurrences of the model responses were divided by their sum to obtain the modeled proportions of responses.

The data from experiment 2, collected by DiNino et al. (2016), were predicted in a similar fashion, using the vocoded VCVs in the considered vocoder conditions as test signals and the unprocessed VCVs as templates. The model’s gammatone filterbank was modified to comprise 20 filters with center frequencies logarithmically spaced between 100 Hz and 8 kHz to take the entire spectral content of the vocoded signals into account. This low-frequency extension was particularly relevant to cover the re-distributed channels in the ApicalSplit condition. In contrast to experiment 1, the experimental stimuli contained no additive noise. Therefore, the temporal patterns of the stimuli and the templates were here directly considered in the model back end, as no “noise alone” pattern could be obtained. Nine iterations of the model simulation were run using newly generated noise-vocoded stimuli in each iteration. As before, internal noise was added and the model response for each iteration was defined as the template showing the largest correlation with the test signal. An internal-noise variance of $\sigma^2_{int,2} = 0.071$ was found to be optimal based on the average recognition scores obtained in the considered conditions, which is in the same range as the variance of 0.05 used in Zaar and Dau (2017) and reflects the relatively low difficulty of the task (cf. Sec. III B). The internal noise was held constant across the considered conditions of experiment 2. For each VCV in each condition, the numbers of occurrences of the model responses were divided by their sum to obtain the modeled proportions of responses.

### III. RESULTS AND ANALYSIS

#### A. Effects of HA signal processing

The grand average consonant recognition scores obtained in the five experimental conditions considered in experiment 1 are shown in Table I. The recognition was at ceiling for the unaided condition (96%), the default HA condition (94%), and the INS condition (92%). In contrast, largely reduced recognition scores were observed in the conditions with NLFC, namely NLFC (55%) and NLFC&INS (56%). The large standard deviations across consonants (36 and 34%, respectively) indicate that the perception of specific consonants was strongly affected by the HA processing while other consonants remained perceptually unaffected. As only the results obtained in the NLFC and NLFC&INS conditions showed substantial perceptual effects of the applied HA processing, the remainder of this section focuses solely on these two conditions.

On average, the model predicted a slightly higher recognition score (59%) than observed in the listeners (55%) for the NLFC condition, whereas it predicted a slightly lower recognition score (51%) than observed in the listeners (56%) for the NLFC&INS condition. To inspect the data more closely in terms of the stimulus-specific recognition and confusion scores, Fig. 2 shows the measured and predicted confusion matrices (CMs) obtained in the NLFC and NLFC&INS conditions. The vertical axes indicate the 12 presented consonants (including the different realizations considered for /s/ and /ʃ/; cf. Sec. II A 1), while the horizontal axes represent the ten consonants provided as response alternatives. The perceptual data (filled gray circles) and the predictions (open red circles) are depicted as circles of different sizes that correspond to the percentage categories shown in the figure’s legend.

In the NLFC condition (left panel of Fig. 2), the listeners exhibited distinct consonant confusions. Most notably, the gray filled circles indicate that /d/ was confused with /b/, /t/ was confused with /k/, /s/ was confused with /ʃ/, /s/ was confused with /ʃ/, /ʃ/ was confused with /ʃ/, and /ʃ/ was confused with /ʃ/. The recognition scores for the mentioned stimuli were thus reduced, with particularly low scores for /s/ and /ʃ/. The model provided convincing predictions of the stimulus-specific recognition scores, as indicated by the good agreement between the red and gray circles on the “diagonal” of the CM (which has two “steps” since two representations of /s/, /ʃ/ were considered as stimuli). Furthermore, the model predicted some of the confusions remarkably well (particularly for /d, s/, s/, s/), although the extent of the confusions was partly underestimated (consistent with the observations in Zaar and Dau, 2017). However, some distinct confusions were not accounted for by the model (/t/ confused with /k/) or predicted to a lesser extent such that they are not visible in Fig. 2. For example, /ʃ/, /ʃ/ were confused with /ʃ/, but the predicted response probabilities for /ʃ/ were just below 7%. Moreover, the model predicted some additional confusions that were not observed in the perceptual data, in particular /s/ was confused with /ʃ/.

The perceptual data obtained in the NLFC&INS condition (right panel of Fig. 2) were largely comparable to the data obtained in the NLFC condition (left panel). However,

| Condition | % correct | Std in % |
|-----------|-----------|----------|
| Unaided   | 95.9      | 8.1      |
| Default   | 93.8      | 7.3      |
| NLFC      | 55.3      | 36.2     |
| INS       | 92.3      | 10.8     |
| NLFC & INS| 56.2      | 34.3     |

TABLE I. Grand average consonant recognition scores measured in experiment 1 for each condition along with the standard deviations across stimuli.
some clear differences can be observed (gray circles), as in the NLFC&INS condition /k/ was confused with /p, f/ and the confusion of /t/ with /k/ observed in the NLFC condition did not occur. Furthermore, /ts/ was not recognized at all in the NLFC condition, but was recognized to some extent in the NLFC&INS condition. The model predictions (red circles) captured these perceptual changes between the NLFC and the NLFC&INS conditions well, apart from the confusion of /k/ with /f/, which was not accounted for by the model.

To evaluate the significance of the agreement between the measured and the predicted stimulus-specific consonant recognition scores (on-diagonal elements of the CMs), a correlation analysis was conducted. Table II summarizes the results, which revealed that the measured and predicted recognition scores were significantly \( p < 0.05 \) correlated across stimuli for both the NLFC \((r = 0.56)\) and the NLFC&INS \((r = 0.67)\) condition.

To also quantify the agreement between the measured and predicted confusions, a correlation analysis of the consonant confusions was performed. For each stimulus, the correlation between the erroneous part of the measured and predicted response patterns (off-diagonal elements of the CMs) was obtained across response alternatives. This analysis was only performed for the stimuli that showed an error of \( P_e > 20\% \) in the perceptual data. Table III shows the results of the confusion correlation analysis, which revealed that the confusions were positively correlated for all considered stimuli, with most correlations being significant. In the NLFC condition, large correlations \((r > 0.88)\) were obtained for /d, s6, s9, S3, S5/ but not for /p, t, ts/, i.e., the confusions were highly correlated for five out of the eight stimuli with \( P_e > 20\% \). In the NLFC&INS condition, large correlations \((r > 0.62)\) were obtained for /k, s6, s9, S3, S5/ but not for /d, ts/, indicating highly correlated confusion patterns for

| Consonant | NLFC | NLFC&INS |
|-----------|------|----------|
|           | \( r \) | \( p \)   | \( r \) | \( p \)   |
| /b/       | -    | -        | -    | -        |
| /g/       | -    | -        | -    | -        |
| /d/       | 0.97 | 0.000    | 0.25 | 0.260    |
| /p/       | 0.16 | 0.344    | -    | -        |
| /k/       | -    | -        | 0.62 | 0.036    |
| /t/       | 0.12 | 0.381    | -    | -        |
| /s6/      | 0.94 | 0.000    | 0.93 | 0.000    |
| /s9/      | 0.97 | 0.000    | 0.97 | 0.000    |
| /S3/      | 0.89 | 0.001    | 0.65 | 0.029    |
| /S5/      | 0.89 | 0.000    | 0.78 | 0.007    |
| /f/       | -    | -        | -    | -        |
| /ts/      | 0.25 | 0.261    | 0.05 | 0.445    |

To evaluate the significance of the agreement between the measured and the predicted stimulus-specific consonant recognition scores (on-diagonal elements of the CMs), a correlation analysis was conducted. Table II summarizes the results, which revealed that the measured and predicted recognition scores were significantly \( p < 0.05 \) correlated across stimuli for both the NLFC \((r = 0.56)\) and the NLFC&INS \((r = 0.67)\) condition.

To also quantify the agreement between the measured and predicted confusions, a correlation analysis of the consonant confusions was performed. For each stimulus, the correlation between the erroneous part of the measured and predicted response patterns (off-diagonal elements of the CMs) was obtained across response alternatives. This analysis was only performed for the stimuli that showed an error of \( P_e > 20\% \) in the perceptual data. Table III shows the results of the confusion correlation analysis, which revealed that the confusions were positively correlated for all considered stimuli, with most correlations being significant. In the NLFC condition, large correlations \((r > 0.88)\) were obtained for /d, s6, s9, S3, S5/ but not for /p, t, ts/, i.e., the confusions were highly correlated for five out of the eight stimuli with \( P_e > 20\% \). In the NLFC&INS condition, large correlations \((r > 0.62)\) were obtained for /k, s6, s9, S3, S5/ but not for /d, ts/, indicating highly correlated confusion patterns for
five out of the seven stimuli with \( P_c > 20\% \). This is consistent with the observations made based on the CMs in Fig. 2, apart from the large confusion correlations found in the two conditions for \( f_s, f_d \), for which the model predicted confusions below 7% (not displayed in Fig. 2). The patterns of predicted confusions were in these cases merely scaled down but qualitatively similar to the measured ones, resulting in large correlations of the confusion patterns.

**B. Effects of CI signal processing**

Table IV shows the grand average measured and predicted consonant recognition scores obtained in the seven experimental conditions of experiment 2 along with the standard deviations across stimuli. As reported by DiNino et al. (2016), the measured recognition scores, including the AllChannels condition, were below ceiling and showed little variability across conditions (73% ± 5%) and a large variability across stimuli (with standard deviations of about 30%). The predicted recognition scores exhibited a similar behavior, albeit with a somewhat smaller variability across stimuli (with standard deviations of about 18.5%).

![Response alternatives](image)

To evaluate the significance of the agreement between the measured and the predicted stimulus-specific consonant recognition scores, a correlation analysis was conducted. Table V summarizes the results, which revealed that the measured and predicted recognition scores (on-diagonal elements of the CMs) were significantly \( p < 0.05 \) correlated for all but the AllChannels condition, capturing the main measured confusions except for \( /th/ \) confused with \( /v/ \). Accordingly, the predicted stimulus-specific recognition scores showed a similar trend as the measured ones, again except for \( /th/ \), for which the model predicted a too high recognition score. The data measured in the BasalZero condition (right panel) also followed the same main trends. However, an additional large perceptual confusion of \( /sh/ \) with \( /s/ \) can be observed along with a reduction in the recognition score for \( /sh/ \). This additional confusion was correctly predicted by the model.

To evaluate the significance of the agreement between the measured and the predicted stimulus-specific consonant recognition scores, a correlation analysis was conducted. Table V summarizes the results, which revealed that the measured and predicted recognition scores (on-diagonal elements of the CMs) were significantly \( p < 0.05 \) correlated for all but the AllChannels condition, capturing the main measured confusions except for \( /th/ \) confused with \( /v/ \). Accordingly, the predicted stimulus-specific recognition scores showed a similar trend as the measured ones, again except for \( /th/ \), for which the model predicted a too high recognition score. The data measured in the BasalZero condition (right panel) also followed the same main trends. However, an additional large perceptual confusion of \( /sh/ \) with \( /s/ \) can be observed along with a reduction in the recognition score for \( /sh/ \). This additional confusion was correctly predicted by the model.
Table VI summarizes the results, which revealed that the confusion correlations for the considered stimuli were very large (mostly above $r = 0.8$) and significant ($p < 0.05$) for the majority of the considered stimuli. However, as observed in the CMs (Figs. 3 and 4), the /th/ confusions were not well predicted by the model (presumably because they originated from a phoneme-frequency effect rather than from the signal characteristics) and the measured and predicted confusions obtained for /b, d/ in the Apical conditions and for /j/ in the BasalZero condition showed either weak correlations or none at all.

**IV. DISCUSSION**

**A. Relation to other studies**

The detrimental effects of NLFC on consonant perception observed in experiment 1 are consistent with the results reported by Schmitt et al. (2016) for HI listeners provided with “too strong” NLFC. The present study, which used a modified version of the speech material from Schmitt et al. (2016), showed similarly strong detrimental effects of NLFC on the recognition of the consonants /s/ and /sh/ (see their Fig. 7). This loss of recognition was shown here to result from a strong confusion of /s/ with /f/, as also discussed in Schmitt et al. (2016). These findings do not contradict studies showing large improvements of high-frequency consonant perception with NLFC in HI listeners (e.g., Glista et al., 2009), as (i) NH listeners were tested in the present study such that no benefit was expected, (ii) strong NLFC settings were used, and (iii) effects

| Condition         | Pearson’s $r$ | $p$-value       | $p$-value       | $p$-value       | $p$-value       |
|-------------------|---------------|-----------------|-----------------|-----------------|-----------------|
|                   |               |                 |                 |                 |                 |
| AllChannels       | 0.43 (0.64)   | 0.051 (0.005)   |                 |                 |                 |
| ApicalZero        | 0.52 (0.75)   | 0.019 (0.001)   |                 |                 |                 |
| ApicalSplit       | 0.45 (0.71)   | 0.040 (0.001)   |                 |                 |                 |
| MiddleZero        | 0.60 (0.79)   | 0.007 (0.000)   |                 |                 |                 |
| MiddleSplit       | 0.61 (0.75)   | 0.006 (0.001)   |                 |                 |                 |
| BasalZero         | 0.31 (0.43)   | 0.122 (0.055)   |                 |                 |                 |
| BasalSplit        | 0.43 (0.66)   | 0.048 (0.007)   |                 |                 |                 |

**IV. DISCUSSION**

**A. Relation to other studies**

The detrimental effects of NLFC on consonant perception observed in experiment 1 are consistent with the results reported by Schmitt et al. (2016) for HI listeners provided with “too strong” NLFC. The present study, which used a modified version of the speech material from Schmitt et al. (2016), showed similarly strong detrimental effects of NLFC on the recognition of the consonants /s/ and /sh/ (see their Fig. 7). This loss of recognition was shown here to result from a strong confusion of /s/ with /f/, as also discussed in Schmitt et al. (2016). These findings do not contradict studies showing large improvements of high-frequency consonant perception with NLFC in HI listeners (e.g., Glista et al., 2009), as (i) NH listeners were tested in the present study such that no benefit was expected, (ii) strong NLFC settings were used, and (iii) effects

| Condition         | Pearson’s $r$ | $p$-value       | $p$-value       | $p$-value       | $p$-value       |
|-------------------|---------------|-----------------|-----------------|-----------------|-----------------|
|                   |               |                 |                 |                 |                 |
| AllChannels       | 0.43 (0.64)   | 0.051 (0.005)   |                 |                 |                 |
| ApicalZero        | 0.52 (0.75)   | 0.019 (0.001)   |                 |                 |                 |
| ApicalSplit       | 0.45 (0.71)   | 0.040 (0.001)   |                 |                 |                 |
| MiddleZero        | 0.60 (0.79)   | 0.007 (0.000)   |                 |                 |                 |
| MiddleSplit       | 0.61 (0.75)   | 0.006 (0.001)   |                 |                 |                 |
| BasalZero         | 0.31 (0.43)   | 0.122 (0.055)   |                 |                 |                 |
| BasalSplit        | 0.43 (0.66)   | 0.048 (0.007)   |                 |                 |                 |
of increasing performance/acclimatization over time (cf. Wolfe et al., 2011) were not considered.

The results from experiment 1 revealed that consonant confusions induced by strong NLFC only occurred within the categories voiced stops, unvoiced stops, and fricatives. Li et al. (2010) and Li et al. (2012) demonstrated that the consonant cues within each of these categories exhibit a similar temporal structure but differ with respect to their spectral energy distributions. The observed effects can therefore be assumed to be caused by spectral changes, resulting in a “too strong” frequency compression applied to the high-frequency consonant cues. The only substantial confusion that did not fall within the above-mentioned categories (/k/ confused with /f/) resulted from combining NLFC with INS, which suppresses sharp onsets and thus produces nonlinear changes over time. The CI processing applied in experiment 2 (DiNino et al., 2016) induced confusions within the categories voiced stops, unvoiced stops, fricatives, and nasals. According to Li et al. (2010) and Li et al. (2012), this again indicates a main perceptual effect of the spectral changes caused by the CI processing.

The present study suggests that the considered model is not limited to conditions of stationary noise but also accounts to a large extent for highly nonlinear signal modifications, implying a versatility that has not been reported so far. The prediction performance was comparable to that reported by Zaar and Dau (2017) for CVs in stationary masking noise in that (i) the predicted stimulus-specific recognition scores were, overall, strongly correlated with the measured recognition scores, (ii) the consonant confusions were mostly well accounted for by the model even though the extent of the confusions was slightly underestimated, and (iii) the model predicted some additional confusions incorrectly (“false alarms”), which mostly fell within perceptually plausible confusion groups. In contrast to Zaar and Dau (2017), effects of masking played a negligible role in the present study as the variability in the perceptual data was mainly induced by changes in the consonant cues of the processed stimuli, resulting in strong confusions. Thus, the results of the present study suggest that the model is sensitive to modifications of consonant cues and the resulting perceptual changes/ambiguities.

B. Comparison to a distance-based modeling approach

Several models have been proposed to account for consonant perception in conditions of additive masking noise (Holube and Kollmeier, 1996; Jürgens and Brand, 2009), using a similar front-end processing as the proposed model (Zaar and Dau, 2017). However, the back-end processing considered by Holube and Kollmeier (1996) and Jürgens and Brand (2009) differed substantially from the proposed model, as it consisted of a DTW algorithm followed by a minimum-distance based decision process, whereas the proposed model additionally considers the noise alone (where applicable), an internal-noise term, and bases its decision on the maximum cross-correlation (for an extensive discussion, see Zaar and Dau, 2017). While the earlier models were either not tested with respect to consonant confusions (Holube and Kollmeier, 1996) or did not account well for consonant confusions in noise (Jürgens and Brand, 2009), they might still predict the consonant morphs considered in the present study. To test this, simulations were also obtained with an alternative version of the model that did not consider the noise alone, the internal-noise term and the correlation metric, but instead based its decision on the minimum cumulative Euclidean distance at the output of the DTW algorithm. This model configuration thus strongly resembles the one proposed by Jürgens and Brand (2009).

An overview of the predictive power of the proposed model and the modified model is given in Table VII. In experiment 1, the modified model showed only slightly less accurate predictions in terms of consonant recognition, as reflected by a minimum average error (MAE) of 6.3% and an average recognition-score correlation \( r_{\text{correct}} \) of 0.58 (proposed model: 4.4% and 0.61, respectively). However, consistent with Jürgens and Brand (2009), the predicted confusion scores obtained with the modified model were inaccurate, as reflected by an average confusion correlation \( r_{\text{conf}} \) of 0.34 (proposed model: 0.63). In experiment 2, the modified model predicted 100% consonant recognition in all stimulus conditions (not shown here). Therefore, the modified model strongly over-predicted the measured overall consonant recognition scores, reflected by a MAE of 27.3% (proposed model: 2.5%). As the modified model predicted 100% recognition for all consonants, there was no variation in the predicted recognition scores across consonants and no consonant confusions were predicted, such that the correlation analyses did not yield any results. Thus, the results indicate that the back end of the model used in the current study (Zaar and Dau, 2017) yields larger predictive power in the considered conditions than the back end considered in the previous models (e.g., Holube and Kollmeier, 1996; Jürgens and Brand, 2009).

C. Limitations of the approach

The model tended to slightly underestimate the extent of the measured consonant confusions and partly predicted additional confusions that were not reflected in the perceptual data. This resulted most likely from a bias induced by either not tested with respect to consonant confusions (Holube and Kollmeier, 1996) or did not account well for consonant confusions in noise (Jürgens and Brand, 2009), they might still predict the consonant morphs considered in the present study. To test this, simulations were also obtained with an alternative version of the model that did not consider the noise alone, the internal-noise term and the correlation metric, but instead based its decision on the minimum cumulative Euclidean distance at the output of the DTW algorithm. This model configuration thus strongly resembles the one proposed by Jürgens and Brand (2009).

An overview of the predictive power of the proposed model and the modified model is given in Table VII. In experiment 1, the modified model showed only slightly less accurate predictions in terms of consonant recognition, as reflected by a minimum average error (MAE) of 6.3% and an average recognition-score correlation \( r_{\text{correct}} \) of 0.58 (proposed model: 4.4% and 0.61, respectively). However, consistent with Jürgens and Brand (2009), the predicted confusion scores obtained with the modified model were inaccurate, as reflected by an average confusion correlation \( r_{\text{conf}} \) of 0.34 (proposed model: 0.63). In experiment 2, the modified model predicted 100% consonant recognition in all stimulus conditions (not shown here). Therefore, the modified model strongly over-predicted the measured overall consonant recognition scores, reflected by a MAE of 27.3% (proposed model: 2.5%). As the modified model predicted 100% recognition for all consonants, there was no variation in the predicted recognition scores across consonants and no consonant confusions were predicted, such that the correlation analyses did not yield any results. Thus, the results indicate that the back end of the model used in the current study (Zaar and Dau, 2017) yields larger predictive power in the considered conditions than the back end considered in the previous models (e.g., Holube and Kollmeier, 1996; Jürgens and Brand, 2009).

| TABLE VII. Global performance evaluation of the proposed model and a distance-based modified version of the model. The MAE describes the mean absolute error in terms of overall consonant recognition, averaged across conditions; \( r_{\text{correct}} \) denotes the Pearson’s correlation between measured and predicted consonant-specific recognition scores (cf. Tables II and V), averaged across conditions; \( r_{\text{conf}} \) represents the confusion correlations (cf. Tables III and VI), averaged across consonants and conditions. The modified model predicted perfect recognition in all conditions of experiment 2 such that the correlation analyses did not yield meaningful results due to the lack of variation. |
|-----------------|-----------------|-----------------|-----------------|
|                 | Experiment 1    | Experiment 2    |
|                 | MAE  | \( r_{\text{correct}} \) | \( r_{\text{conf}} \) | MAE   | \( r_{\text{correct}} \) | \( r_{\text{conf}} \) |
| Proposed model  | 4.4% | 0.61 | 0.63 | 2.5% | 0.48 | 0.64 |
| Modified model  | 6.3% | 0.58 | 0.34 | 27.3% | -   | -   |
similarities/dissimilarities in the vowel portions of the CV/VCV stimuli and templates, which are only partly related to the consonant percept. However, a separation of the signals into consonant and vowel portions is not feasible, particularly for speech tokens containing voiced consonants, which rely on formant transitions in the adjacent portions of the accompanying vowels.

Furthermore, the model does not take any linguistic processing (e.g., biases) into account, which may affect the perceptual results to some extent despite the nonsense nature of the stimuli. For example, the consistent perceptual morph of /th/ to /v/ in experiment 2 presumably resulted from a perceptual bias induced by the large phoneme frequency of /v/ and the low phoneme frequency of /th/ to /v/ in the English language. This cannot be accounted for by the model, which evaluates solely the similarity of the signals.

D. Perspectives

An important extension of the model would be to include aspects of hearing impairment, such as elevated audiometric thresholds, reduced frequency selectivity, loss of compression and other supra-threshold deficits (cf. Jürgens et al., 2014; Jepsen et al., 2014). The results of the present study suggest that, if a version of the model that can account for consonant perception in unaided HI listeners was established, the effects of hearing-instrument compensation strategies might be well-represented in the model predictions. Thus, the model could provide guidance regarding HA fitting, e.g., by suggesting specific fitting parameters based on a listener’s auditory profile. Furthermore, it might become feasible to predict effects of acclimatization in aided HI listeners and CI users by assuming different amounts of a priori knowledge about the processing strategy in the model back end.

V. SUMMARY AND CONCLUSION

The present study evaluated the predictive power of the model of Zaar and Dau (2017) regarding effects of HA and CI signal processing on consonant perception. Experiment 1 considered consonant perception in NH listeners after HA processing in terms of nonlinear frequency compression and impulse-noise suppression using CVs. Experiment 2 considered consonant perception in NH listeners after CI processing with different simulations of poor electrode-neuron interfaces using VCVs. The model was shown to account for most perceptual effects observed in the data from experiment 1. In particular, the stimulus-specific predicted recognition scores were significantly correlated with the measured ones, as well as most of the stimulus-specific confusion patterns. Furthermore, the model accounted to a large extent for the data from experiment 2, i.e., for the effects of the CI signal processing on consonant perception. Specifically, the simulated stimulus-specific recognition scores were significantly correlated with the measured ones in most conditions. Moreover, the vast majority of the stimulus-specific predicted confusion patterns were highly significantly correlated with the perceptual data.

The results indicate that the presented modeling approach, which was earlier shown to account for consonant recognition and confusions obtained with CVs in stationary noise (Zaar and Dau, 2017), also accounts for supra-threshold effects of hearing-instrument signal processing on consonant perception reasonably well. This suggests a large potential of the model for evaluating and adjusting such processing schemes, in particular when extended to account for individual hearing impairment.

ACKNOWLEDGMENTS

This research was funded with support from the European Commission under Contract No. FP7-PEOPLE-2011-290000. The authors wish to thank Leonid Litvak for providing the vocoder simulation code used in experiment 2.
Litvak, L. M., Spahr, A. J., Saoji, A. A., and Fridman, G. Y. (2007). "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners," J. Acoust. Soc. Am. 122(2), 982–991.

Miller, G. A., and Licklider, J. C. R. (1950). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. 22(2), 167–173.

Miller, G. A., and Nicely, P. E. (1955). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. 27(2), 338–352.

Nielsen, J. B., and Dau, T. (2009). "Development of a Danish speech intelligibility test," Int. J. Audiol. 48(10), 729–741.

Nielsen, J. B., and Dau, T. (2011). "The Danish hearing in noise test," Int. J. Audiol. 50(3), 202–208.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. 95(2), 1085–1099.

Pavlovic, C. V. (1987). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. 82(2), 413–422.

Phatak, S. A., and Allen, J. B. (2007). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. 121(4), 2312–2326.

Phatak, S. A., Lovitt, A., and Allen, J. B. (2008). "Consonant confusions in white noise," J. Acoust. Soc. Am. 124(2), 1220–1233.

Phatak, S. A., Yoon, Y.-S., Gooler, D. M., and Allen, J. B. (2009). "Consonant recognition loss in hearing impaired listeners," J. Acoust. Soc. Am. 126(5), 2683–2694.

Sakoe, H., and Chiba, S. (1978). "Dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. Acoust. Speech Signal Process. 26, 43–49.

Scheidiger, C., Allen, J. B., and Dau, T. (2017). "Assessing the efficacy of hearing-aid amplification using a phoneme test," J. Acoust. Soc. Am. 141(3), 1739–1748.

Schmitt, N., Winkler, A., Boretzki, M., and Holube, I. (2016). "A phoneme perception test method for high-frequency hearing aid fitting," J. Am. Acad. Audiol. 27(5), 367–379.

Simpson, A., Hensbach, A. A., and McDermott, H. J. (2005). "Improvements in speech perception with an experimental nonlinear frequency compression hearing device," Int. J. Audiol. 44(5), 281–292.

Stickney, G. S., Loizou, P. C., Mishra, L. N., Assmann, P. F., Shannon, R. V., and Opie, J. M. (2006). "Effects of electrode design and configuration on channel interactions," Hear. Res. 211, 33–45.

Trevino, A., and Allen, J. B. (2013). "Within-consonant perceptual differences in the hearing impaired ear," J. Acoust. Soc. Am. 134(1), 607–617.

Wagener, K., Jovassen, J. L., and Ardenkjær, R. (2003). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol. 42, 10–17.

Wang, M. D., and Bilger, R. C. (1973). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. 54(5), 1248–1266.

Warren, R. M. (1970). "Perceptual restoration of missing speech sounds," Science 167, 392–393.

White, M. W., Merzenich, M. M., and Gardi, J. N. (1984). "Multichannel cochlear implants. Channel interactions and processor design," Arch. Otolaryngol. 110(8), 493–501.

Wolfe, J., John, A., Schafer, E., Nyffeler, M., Boretzki, M., Caraway, T., and Hudson, M. (2011). "Long-term effects of non-linear frequency compression for children with moderate hearing loss," Int. J. Audiol. 50(6), 396–404.

Zaar, J., and Dau, T. (2015). "Sources of variability in consonant perception of normal-hearing listeners," J. Acoust. Soc. Am. 138(3), 1253–1267.

Zaar, J., and Dau, T. (2017). "Predicting consonant recognition and confusions in normal-hearing listeners," J. Acoust. Soc. Am. 141(2), 1051–1064.