High-Frequency Sensorineural Hearing Loss Alters Cue-Weighting Strategies for Discriminating Stop Consonants in Noise

Léo Varnet, Chloé Langlet, Christian Lorenzi, Diane S. Lazard, Christophe Micheyl

Original Article

High-Frequency Sensorineural Hearing Loss Alters Cue-Weighting Strategies for Discriminating Stop Consonants in Noise

Trends in Hearing
Volume 23: 1–18

DOI: 10.1177/2331216519886707
journals.sagepub.com/home/tia
SAGE

Léo Varnet[1], Chloé Langlet[1], Christian Lorenzi[1], Diane S. Lazard[2], and Christophe Micheyl[3]

Abstract

There is increasing evidence that hearing-impaired (HI) individuals do not use the same listening strategies as normal-hearing (NH) individuals, even when wearing optimally fitted hearing aids. In this perspective, better characterization of individual perceptual strategies is an important step toward designing more effective speech-processing algorithms. Here, we describe two complementary approaches for (a) revealing the acoustic cues used by a participant in a /d/-/g/ categorization task in noise and (b) measuring the relative contributions of these cues to decision. These two approaches involve natural speech recordings altered by the addition of a "bump noise." The bumps were narrowband bursts of noise localized on the spectrotemporal locations of the acoustic cues, allowing the experimenter to manipulate the consonant percept. The cue-weighting strategies were estimated for three groups of participants: 17 NH listeners, 18 HI listeners with high-frequency loss, and 15 HI listeners with flat loss. HI participants were provided with individual frequency-dependent amplification to compensate for their hearing loss. Although all listeners relied more heavily on the high-frequency cue than on the low-frequency cue, an important variability was observed in the individual weights, mostly explained by differences in internal noise. Individuals with high-frequency loss relied slightly less heavily on the high-frequency cue relative to the low-frequency cue, compared with NH individuals, suggesting a possible influence of supra-threshold deficits on cue-weighting strategies. Altogether, these results suggest a need for individually tailored speech-in-noise processing in hearing aids, if more effective speech discriminability in noise is to be achieved.

Keywords

hearing aids, sensorineural hearing loss, speech perception

Received 15 November 2018; Revised 30 September 2019; accepted 11 October 2019

Introduction

As for any communication device, the decoding of speech by the human auditory system relies on a "code" associating a physical input, the speech sound, with some linguistic representations, such as syllables. This acoustic-linguistic conversion requires the detection of specific features present in the incoming signal, termed "acoustic cues," which are associated with particular phonetic segments (Allen, 1994). This is not a one-to-one relationship, however, as phonetic distinctions may rely on the integration of multiple cues (Clayards, 2018; Delattre, 1968). Speech sounds are highly redundant in their acoustical content, so several correlated acoustic cues are often available to distinguish between two syllables, ensuring high flexibility and robustness to the human speech perception system (e.g., Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). On the other hand, poor detection of the cues or suboptimal processing of the information that they convey, as may happen in individuals with mild or severe hearing loss, can lead to a decrease in intelligibility (Phatak, Yoon, Gooler, & Allen, 2009).

[1]Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, Université Paris Sciences et Lettres, CNRS, Paris, France
[2]Institut Arthur Vernes, Paris, France
[3]Starkey Hearing Technologies, Eden Prairie, MN, USA

Corresponding Author:
Léo Varnet, Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, Université Paris Sciences et Lettres, CNRS, 75005 Paris, France.
Email: leo.varnet@ens.fr

Understanding which acoustic cues human listeners rely on to discriminate speech sounds, how they use and combine such cues, and how these perceptual strategies are impacted by hearing loss, are important steps toward designing more effective speech-processing algorithms for hearing-impaired (HI) or normal-hearing (NH) individuals. The goal of this study was to examine these questions by focusing on the example of voiced stop consonant categorization in noise by individuals with or without hearing loss.

The earliest psychoacoustic studies of voiced stop consonant perception relied on the use of synthetic speech continua to demonstrate that varying the second (Liberman, Delattre, Cooper, & Gerstman, 1954) and third (Mann, 1980) formant onsets (F2 and F3 onsets) affects the perception of the stimulus as an instance of /d/ or /g/. This result has been replicated many times since, confirming that F2/F3 onsets are a primary cue for this task (see, e.g., Delattre, 1968; Holt, 2005; Viswanathan, Magnuson, & Fowler, 2010). In addition, several researchers have suggested a secondary role of F1 onset in place perception. In an early exploratory study, Delattre, Liberman, and Cooper (1955) noticed that, when the primary F2 cue was ambiguous (onset midway between /d/ and /g/), the phonetic decision was driven by the height of the F1 onset. A closer examination of natural recordings reveals a small ($\sim$100 Hz) but very consistent difference between the F1 onset in /da/ and /ga/ (as can be seen, e.g., in Mann, 1980; Summers & Leek, 1997; Turner, Fabry, Barrett, & Horwitz, 1992; Varnet, Meunier, Trollé, & Hoen, 2016), which may be used as a cue by the listener. Furthermore, when the first formant of a stop consonant is removed by filtering, the stimulus is less well identified (Summers & Leek, 1997), whereas such removal does not affect the recognition scores for artificial stimuli where F1 characteristics are held constant (Dorman, Lindholm, & Hannley, 1985). Summers and Leek also reported that this effect of F1 suppression is particularly strong when the speech stimuli are presented in noise. Although these pioneering studies using synthetic or artificially modified stimuli had a major impact on speech perception research, a recurrent criticism of this methodology is that the resulting sounds are very unnatural and, therefore, that the results may not be generalizable to everyday speech perception (Hazan & Rosen, 1991; Li, Menon, & Allen, 2010).

Recently, Varnet, Knoblauch, Meunier, and Hoen, (2013) developed a new psychophysical reverse correlation (revcorr) method to uncover perceptually relevant acoustic cues for consonant discrimination using natural speech stimuli (see also Brimijoin, Akeroyd, Tilbury, & Porr, 2013; Mandel, Yoho, & Healy, 2016 for other examples of speech auditory revcorr experiments). They had participants listen to a series of speech utterances embedded in white noise with a low signal-to-noise ratio (SNR; -10.7 dB on average), and they recorded the effect of the noise sample upon the syllable categorization on a trial-by-trial basis. Then, they related the spectrotemporal content of the noise in each trial with the corresponding response of the participant, using a generalized linear model (GLM). Each time–frequency bin in the stimulus spectrogram was associated with one weight in the phonetic decision, resulting in a spectrotemporal matrix of weights termed an auditory classification image (ACI). When calculated in the context of a /da/-/ga/ categorization task in noise, the ACIs of NH listeners consistently show a strong cluster of weights in the spectrotemporal region of the second- and third-formant (F2–F3) onsets, and a weaker set of weights in the first-formant (F1) onset region (Varnet, Knoblauch, Serniclaes, Meunier, & Hoen, 2015; Varnet, Wang, Peter, Meunier, & Hoen, 2015). These results confirm that this particular phonetic decision mainly relies on the detection of a primary F2–F3 cue and a secondary F1 cue.

However, another set of studies based on natural speech sounds has instead highlighted the key role of prevocalic bursts in stop consonant perception (Kapoor & Allen, 2012; Li et al., 2010; Li & Allen, 2011; Mackersie, 2007; Ohde & Stevens, 1983; Summers & Leek, 1997). Kapoor and Allen suggest that these bursts constitute a primary cue for correctly identifying /t/, /d/, /g/, and /b/. At first sight, this result seems in direct contradiction with the aforementioned ACI experiments. To reconcile these observations with those of Varnet et al., one must consider the effect of background noise level. When the SNR is low, as in the case of the ACI experiment (speech in white noise with SNR ≈ -10 dB), the burst becomes far less audible than the formants (Régnier & Allen, 2008; Summers & Leek, 1997), especially in voiced stop consonants (Li et al., 2010). Kapoor and Allen compared recognition scores for natural utterances of stop consonants in noise, either unmodified, or with the burst feature manually removed. The presence of the burst improved intelligibility for the highest SNR (≥ -6 dB) but not when SNR $= -12$ dB (Kapoor & Allen, 2012). Therefore, it seems plausible that the burst was the predominant cue for speech perception in quiet while, for low SNRs where the burst cue was not audible, the auditory system switched to the use of formant information.

In the same vein, Serniclaes and Arrouas (1995) have studied the /dɔ/-/tɔ/ contrast, which involves at least three cues. The primary cue is the voice onset time (VOT), the period of time between the release of the tongue and the onset of the vocal fold vibrations. Other cues available to the listener include the fundamental frequency and formant trajectories at the onset of the consonant. The authors have shown that, in the

absence of background noise, listeners rely on the VOT cue only. However, in the presence of a background noise, the VOT cue becomes less reliable and listeners switch to the use of the transition cues. More generally, there seems to be a dichotomy between primary cues, which strongly affect categorization, and secondary cues, which have a lesser influence on perception or are used only when the primary cues are removed or degraded (Li et al., 2010; Varnet, Meunier, & Hoen, 2016).

The previous paragraphs focused on short-term adaptations. However, cue-weight changes may also occur on longer scales when difficulties arise from the listener himself instead of his or her immediate acoustic environment. For example, HI listeners may adapt their speech-perception strategies depending on their specific hearing loss profile. In particular, it has been suggested that some listeners with high-frequency hearing loss have learned to rely more heavily on low-frequency than on high-frequency cues, the latter being either less audible or more distorted for these listeners (Moore & Vinay, 2009; Seldran et al., 2011; Turner & Brus, 2001).

The most straightforward approach for investigating phoneme perception by HI listeners is to examine their patterns of error in phoneme-identification tasks. In addition to the overall performance level, the distribution of confusions is informative about the type of errors made. Confusion matrices have been measured for HI listeners both in quiet and at very low noise levels (Bilger & Wang, 1976; Dubno, Dirks, & Langhofer, 1982; Owens, 1978) or with various SNRs (Phatak et al., 2009; Trevino & Allen, 2013). The distribution of HI participants' answers reveals specific patterns of confusions, different from those of NH listeners (Scheidiger & Allen, 2013; Scheidiger, Allen, & Dau, 2017; Trevino & Allen, 2013). Place of articulation errors are most frequent in HI listeners, regardless of audiometric configuration (Bilger & Wang, 1976; Dubno et al., 1982; Owens, 1978; Turner & Brus, 2001). In addition to having lower overall performance than NH listeners, HI listeners also show great variability (Phatak et al., 2009) when compared with a control group with no hearing deficits (Phatak & Allen, 2007). Some of this variability can be accounted for based on audiometric thresholds (Bilger & Wang, 1976). In particular, scores obtained by listeners with a high-frequency loss are generally poorer than those obtained by listeners with a flat loss (Dubno et al., 1982), notably for plosive consonants (Phatak et al., 2009). However, speech audibility is generally a poor predictor of intelligibility in HI listeners (Glasberg & Moore, 1989; Seldran et al., 2011). A recurrent finding has been that the restoration of audibility through the use of amplification provides only a limited benefit for intelligibility (Abavisani & Allen, 2017; Hogan & Turner, 1998; Plomp, 1978; Scheidiger & Allen, 2013). Furthermore, for hearing aid users, interindividual differences in phoneme recognition performance are not well predicted by absolute hearing thresholds (Bernstein et al., 2016; Humes, 2007), therefore suggesting the existence of additional suprathreshold deficits (Lesica, 2018; Plomp, 1978).

A few studies have tried to relate the intelligibility of phoneme utterances to the audibility of acoustic cues in a particular instance of the phoneme. Using such an approach, Turner and Robb (1987) and Turner and Brus (2001) have shown that, contrary to NH listeners, whose scores were directly related to audibility of acoustic cues, HI listeners were unable to make efficient use of acoustic cues even when these cues were presented at suprathreshold levels. These findings are consistent with the notion that speech recognition depends on auditory deficits beyond the mere loss of absolute sensitivity that usually characterizes hearing loss. In the same vein, Turner et al. (1992) showed that, even if listeners with and without hearing losses have very similar psychometric functions for consonant-in-noise *detection*, the former show poorer psychometric functions for consonant *identification* in noise. This inability to make use of available information suggests that HI listeners may apply different listening strategies, and rely on different cues, than NH listeners. As noted by Trevino and Allen (2013), the fact that utterances that are better identified by NH participants are not necessarily the more robust for HI participants is also evidence for a different use of available acoustic cues.

Correlational methods have been extensively used to explore the strategies of listeners in different auditory categorization tasks. Once an estimate of NH listeners weighting strategy has been obtained, some researchers tried to compare it with the data of HI listeners performing the same task. Early examples come from two studies by Doherty and Lutfi. These researchers measured weighting strategies for both NH and HI listeners on a nonspeech, level-discrimination task, either for a complex tone (Doherty & Lutfi, 1996) or for one single component of this complex (Doherty & Lutfi, 1999) and showed that HI listeners were more sensitive to frequencies associated with their cochlear damage. The authors concluded that these participants may put more weight on the information within the region of their hearing loss to compensate for the degraded sensory information in those regions.

A different kind of frequency weighting-function, close to the "frequency-importance function" of the Speech Intelligibility Index (American National Standards Institute, 1997), has been derived in the case of speech sentence comprehension in broadband noise for NH listeners (Calandruccio & Doherty, 2007) and HI listeners (Calandruccio & Doherty, 2008). The latter group was tested both with and without hearing aid correction (using a NAL-R fitting algorithm). Contrary to the

previous experiment by Doherty and Lutfi, the regression variable was not the content of the target but the SNR in each frequency band. Therefore, each weight reflects the difference in intelligibility when a specific band becomes masked. For both conditions, the authors observed that the weighting of the 1787 to 2807 Hz frequency band was less for HI than NH individuals. Calandruccio and Doherty interpreted this as reflecting a different use of the formant transition cues by HI listeners. In addition, when not wearing hearing aids, HI listeners put more weight on the high frequency band (2807–11000 Hz). Similar conclusions were reached by Gilbert, Micheyl, Berger-Vachon, and Collet (2002) using a very similar experimental procedure on NH listeners and HI listeners without hearing aids.

These studies, using sentence stimuli, provide some insight into the perceptual frequency-weighting strategies of NH and HI listeners for speech perception in noise. However, they provide only limited insight into the difficulties of these individuals with specific phonetic features such as place of articulation (Bilger & Wang, 1976; Dubno et al., 1982; Owens, 1978; Turner & Brus, 2001). A possible solution to identify the acoustic cues that are used by these listeners to discriminate specific phonemes is to apply the same correlational approach at the "microscopic" level by artificially removing or enhancing the cues and observing how this affects intelligibility. Pittman and Stelmachowicz (2000) used four vowel-fricative stimuli, divided into three temporal segments corresponding roughly to the vowel, the formantic transition, and the fricative segment. Each of these segments was presented at a randomly chosen level. Then, these levels were correlated with recognition scores on a trial-by-trial basis. The researchers compared the results of NH and HI participants in this task and observed that all groups weighted the fricative segment more heavily for /s/ and /ʃ/, and all three segments equally for /f/ and /θ/, although small quantitative differences were observed in the weightings. These conclusions were confirmed in a second experiment using the same approach with different stimuli (Pittman, Stelmachowicz, Lewis, & Hoover, 2002).

The present study aimed at evaluating quantitatively the relative weightings of two cues involved in a phonetic decision, for HI individuals with different audiometric configurations. This was motivated by previous reports of changes in listening strategy following cochlear damage, despite restored audibility through linear amplification. A novel experimental paradigm derived from the ACI methodology was used, allowing for more controlled and more precise manipulations of the various acoustic cues involved in the perception of a given phonetic contrast. A /d/-/g/ categorization task was chosen because place of articulation contrasts is known to be particularly challenging for HI listeners, even when they are provided with a frequency-dependent amplification. First, a preliminary pilot experiment based on a revcorr approach very similar to that of Varnet, Knoblauch, et al. (2015) aimed at determining the frequency location of the formant acoustic cues for four utterances of "Alda," "Alga," "Arga," and "Arda." Then, in a second experiment, we actively manipulated these acoustic cues to bias the perception of the participants. The weights on the primary and secondary cues, estimated through a GLM model of the phonetic decision, were obtained for each individual and each group and compared between NH listeners and HI listeners with a high-frequency loss corrected with a simulated hearing aid. An additional group of HI listeners with a relatively flat audiometric profile in the region of the cues was included in an attempt to find evidence for a relationship between pure-tone thresholds and cue-weighting strategy.

Pilot Experiment

To confirm that two formant onset cues were involved in the da/ga categorization task and determine their exact frequency locations, we conducted a revcorr experiment very similar to the ACI experiment described in the "Introduction" section (Varnet, Knoblauch, et al., 2015) but based on a different type of noise called "bump noise." This was done in order to restrict the number of parameters (degrees of freedom) in the description of the noise and thus reduce the duration of the experiment. Such "dimensional noise" approaches, where the noise is applied only to one dimension of interest of the stimuli, have already been successfully used in previous visual revcorr studies (Kurki & Eckstein, 2014; Kurki, Saarinen, & Hyvärinen, 2014; Li, Klein, & Levi, 2006).

Stimuli, Participants, and Procedure

Seven NH participants were asked to listen to a series of noisy bisyllabic pseudowords (/alda/, /alga/, /aʁda/, or/aʁga/), in random order, and to categorize the second syllable as a "da" or a "ga." The four target stimuli were the same as those used in previous experiments (Varnet, Knoblauch, et al., 2015; Varnet, Meunier, Trollé, et al., 2016; Varnet, Wang, et al., 2015) with the only refinement being that their pitch contours were made similar in Praat in order to avoid possible stimuli-specific strategies based on subtle differences in f0 trajectories. The spectrograms of the stimuli are shown in Figure 1.

As a first step, the participant's SNR threshold was determined through an adaptive 2-down 1-up staircase procedure (mean of three measurements) targeting a performance level of 70.7% correct. During this stage, stimuli were presented in a white noise masker.

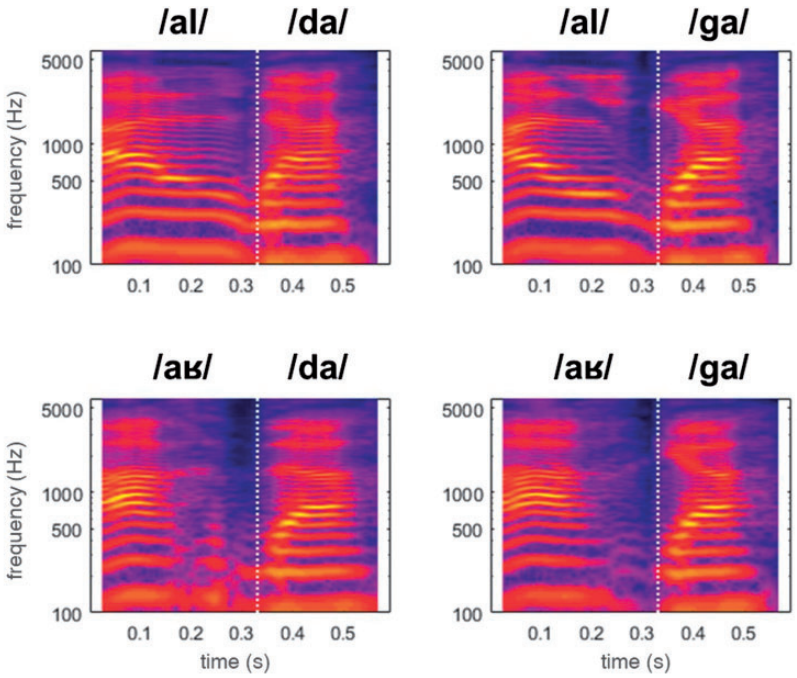

Figure 1. Spectrograms of the four stimuli used in this study (frequency axis displayed with a logarithmic scale). The white dotted line at $t = 0.33$ s marks the boundary between the two syllables and the temporal position of the bumps in the two experiments.

In the second part of the experiment, 1,000 stimuli were presented to each participants at his or her individual SNR threshold level. The masker was a bump noise comprising four bumps at random locations, as described later. The experiment was divided into blocks of 200 stimuli separated with breaks, to limit mental fatigue.

Participants sat in front of a computer screen in a sound-proof cabin and responded by using the mouse or the keyboard. All stimuli were presented diotically at 70 dB SPL through Sennheiser HD600 headphones and an external audioengine D3 digital-to-analog audio converter (Austin, TX).

The experiments were run under MATLAB R2016b (The Mathworks, Natick, MA) using the AFC toolbox (Ewert, 2013).

Bump Noise

The bump noise is designed to manipulate specific spectrotemporal regions of the speech target. It is very similar to the "bubble noise" of Mandel et al. (2016), except that the former is composed of Gaussian bumps superimposed on white noise, while the latter consists of Gaussian holes in a speech-shaped noise.

In the pilot experiment, the bump noise added to the targets was composed of four Gaussian bumps (identified by $i \in \{1. \ . \ . \ N\}$ with $N = 4$), temporally aligned with the onset of the second syllable in the targets ($t_i = 330$ ms for all i). The center frequencies of the bumps, f_i, were chosen randomly at the beginning of each trial. Two center frequencies were drawn from the interval [50 Hz, 1000 Hz] and two others from the interval [1000 Hz, 5186.6 Hz] from a uniform distribution on the $\mathrm{ERB_N}$ scale (Moore, 2005). The distance between two center frequencies was at least 2 $\mathrm{ERB_N}$. The widths (corresponding to the standard deviations) of the Gaussian bumps were $\sigma_t = 30$ ms on the time axis and $\sigma_f = 1$ $\mathrm{ERB_N}$ on the frequency axis. The scaling factor A_i controlling the amplitude of the bump relative to the background noise, and therefore expressed in units of baseline noise level, was the same for the four bumps ($A_i = 7$). This value was chosen empirically to be sufficiently large for the bumps to measurably influence the decision of the observer, yet sufficiently small to avoid perceptual segregation of the bumps from the remainder of the stimulus, which could have interfered with the measurements.

The spectrotemporal envelope of the bump noise is described by Equation 1.

$$B(f,t) = 1 + \sum_{i=1}^{N} A_i \cdot \exp\left[-\frac{(t-t_i)^2}{\sigma_t^2} - \frac{(E(f)-E(f_i))^2}{\sigma_f^2}\right] \tag{1}$$

with $E(f) = q \cdot \log\left(1 + \frac{f}{l \cdot q}\right)$, $l = 24.7$ and $q = 9.265$ (Hohmann, 2002).

This "ideal" template was multiplied by the spectrogram of a white noise to obtain the spectrogram of the masker (white noise plus bumps).

Results

Overall, participants obtained 60.2% correct (±5.8% *SD*) on average in the pilot experiment (using bump noises). The SNR level at which they performed the task corresponded to a theoretical 70.7% correct recognition in white noise (performance level targeted by the initial staircase, see Table 1). Therefore, the masking effect due to the addition of four random bumps on the onset of the syllable can be estimated as corresponding to approximately a 10 percentage-point change in overall performance (i.e., participants made an additional 10% errors when targets were presented in bump noise instead of white noise).

Figure 2 shows the frequency distribution of the bumps across trials on which the participants responded "da" (red line) or "ga" (blue line). This percentage representation was preferred over a simple count of the "da" and "ga" bumps at each frequency because it corrects for a possible bias of the participants toward one response.

Figure 2 confirms that there were two main critical regions where noise influences the decision of the listener: a high-frequency region (between 1400 and 2700 Hz) corresponding to the F2 and F3 onsets, which will be referred to as "HF cue" hereafter, and a weaker low-frequency region (between 350 and 700 Hz) corresponding to the F1 onset, which will be referred to as "LF cue" (shaded regions in Figure 2). These results allowed us to determine the center frequencies for which the bumps impacted the perceptual phonetic categorization of the stimuli the most, on average. The bump center frequencies that were retained, to be used in the main experiment, were the following: 578 Hz ("da"-percept-inducing bumps on the F1 onset), 1500 Hz and 2641 Hz ("da"-percept-inducing bumps on the F2/F3 onsets), 390 Hz ("ga"-percept-inducing bumps on the F1 onset), and 1975 Hz and 2125 Hz ("ga"-percept-inducing bumps on the F2/F3 onsets). Note that the choice of the exact frequency values was somewhat arbitrary as the distributions in Figure 2 appear to be quite noisy. However, our main focus in this experiment was to ensure that the bump noise can actively bias the phonetic decision of the participant toward one response or the other, which turned out to be the case.

Parametric Bump Noise Experiment

The main aim of this study was not to identify the acoustic cues involved in the da/ga categorization in noise, already known from previous studies (e.g., Varnet, Knoblauch, et al., 2015), but rather to evaluate quantitatively the relative contributions of these cues to the phonetic decision. Accordingly, the main experiment was designed to measure the sensitivity of the listener to the earlier defined cues. By varying parametrically the amplitude of the bumps in the masker from a "ga"-percept-inducing bump noise to a "da"-percept-inducing bump noise

Table 1. Summary of the Characteristics of the Three Groups.

Group name	*N*	Age (years)	Hearing aid experience (months)	SNR (dB)
NH	17	27.4 ± 3.6 *SD*	–	−10.4 ± 2.0 *SD*
HI–HF	18	64.3 ± 6.3 *SD*	36.1 ± 34.3 *SD*	−6.7 ± 2.3 *SD*
HI–flat	15	62.9 ± 6.4 *SD*	20.8 ± 24.5 *SD*	−6.0 ± 2.5 *SD*

Note. SNR = signal-to-noise ratio; NH = normal-hearing; HI = hearing-impaired; HF = high-frequency loss; flat = flat loss.

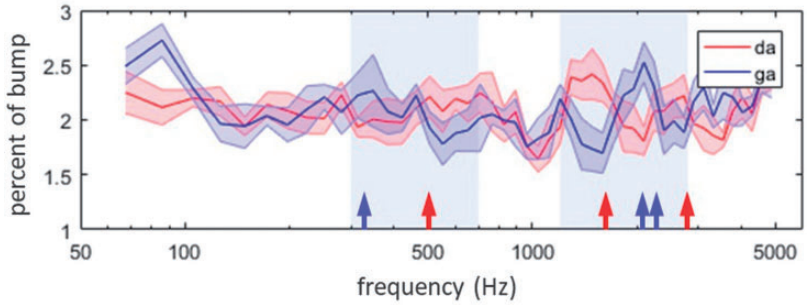

Figure 2. Mean distributions of the bumps yielding a "da" (red line) or a "ga" (blue line) response as a function of their frequency (±0.5 *SD*). The two shaded areas represent the regions of the HF and LF cues. Arrows mark the approximate locations of F1, F2, and F3 onsets for "da" (red arrow) and "ga" (blue arrow).

(see later) and measuring the proportion of confusions in the labeling of the masked speech stimuli, we were able to estimate psychometric functions corresponding to each of the two cues.

Participants

Three groups of listeners participated in the main experiment. All participants were native speakers of French. The first consisted of 17 young adults (age = 27.4 years ± 3.6 *SD*), all with normal audiometric thresholds (≤20 dB HL) for octave frequencies between 125 and 8000 Hz. This group will be referred to as the NH group. The second and third groups consisted of older listeners with sensorineural hearing loss profiles in the right ear. The 18 individuals in the high-frequency loss (HI-HF) group had moderate to severe loss in the 1000 Hz to 8000 Hz region (audiometric thresholds > 30 dB HL) but normal or near-normal between 125 and 750 Hz (thresholds ≤20 dB HL). Their age ranged from 55 to 73 years (mean = 64.3 years ± 6.3 *SD*). The 15 individuals in the flat loss (HI-flat) group had moderate and flat or quasi-flat loss in the 500 Hz to 4000 Hz region (thresholds between 20 dB HL and 50 dB HL with a maximum difference of 15 dB). Their age ranged from 51 to 72 years (mean = 62.9 years ± 6.4 *SD*). Across the two HI groups, 29 participants were current users of hearing aids and 4 had no or very little (<1 month) previous experience with hearing aids. All hearing losses were of sensory origin, as confirmed by the absence of air-bone gaps in the audiometric thresholds. Although the experiment was only conducted on the right ear, we made sure that the hearing losses were broadly symmetrical (between-ear difference of maximum 15 dB). We excluded from this study all listeners suffering from tinnitus or Ménière's disease, or having any psychiatric disorders.

Figure 3 shows individual and mean right ear audiograms for the three groups. A summary of the characteristics of the three groups is provided in Table 1.

All listeners were fully informed about the goal of the study, provided written consent, and received financial compensation for their participation. The study received the approval of the Ethical Committee CPP Ile de France III with the ID RCB: 2016-A0176901769-42.

Stimuli and Procedure

In the main experiment, all stimuli were presented monaurally to the right ear at 70 dB SPL. For HI participants, the sounds were amplified in a frequency-dependent manner depending on their pure-tone audiogram, using the NAL-R formula (Byrne & Dillon, 1986; Palmer & Lindley, 2002).

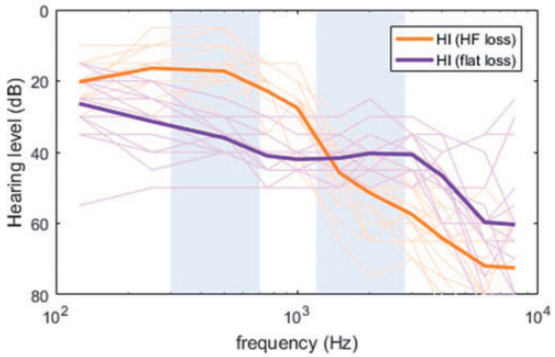

Figure 3. Individual and mean audiometric thresholds for the listeners with HF (orange line) or flat (indigo line) loss for the right ear. The two shaded areas represent the regions of the HF and LF cues.

The target sounds were the same as in the pilot experiment, and the general procedure was largely similar: First, the individual SNR threshold for 70.7% correct categorization in white noise was determined by means of an adaptive 2-down 1-up staircase procedure. Then, the SNR was fixed at this level for the second phase of the experiment.

The purpose of this experiment was not to find the location of the acoustic cues, as in the pilot study, but rather to measure the sensitivity of the listener to predefined cues. Accordingly, the bump noises used here were slightly different from those described earlier. As before, they were generated according to Equation 1, with $t_i = 330\ ms$, $\sigma_t = 30\ ms$, and $\sigma_f = 1\ ERB$. However, they were composed of six bumps with fixed frequency positions (three formant onset frequencies for the "da"-percept-inducing bump and three formant onset frequencies for the "ga"-percept-inducing bump). The four higher frequency bumps (at 1500, 1975, 2641, and 2125 Hz) corresponded to the primary HF cue on the F2/F3 onsets, while the two lower bumps (at 390 and 578 Hz) corresponded to the secondary LF cue on the F1 onset (see Table 2). The two ga-percept-inducing HF bumps were relatively close (1975 Hz and 2125 Hz) and therefore overlapped to some extent. This was not an issue for the current investigation, however, as the F2 and F3 onsets were considered as a single cue (in line with the ACIs in Varnet, Knoblauch, et al., 2015 which show a single cluster of weights between the two formants).

Contrary to the pilot experiment, the bump amplitudes, A_i, were not equal from one trial to another and between bumps. We created a two-dimensional continuum of bump noise profiles $B(f,\ t)$ by varying orthogonally the amplitudes of the HF cue bumps and those of the LF cue bumps from 0 to 5. There were five levels for each of the two cues, totaling 25 bump noise profiles, illustrated in Figure 4. The top left condition corresponds to the most /da/-percept-inducing configuration

of the bumps (level of all da-bumps set to 5, level of all ga-bumps set to 0), and the bottom right condition to the most /da/-percept-inducing configuration (level of all da-bumps set to 0, level of all ga-bumps set to 5). These bump noises were superimposed with one of the four possible targets in a full-factorial design. Note that, because of the presence of bumps in each condition, it is very likely that the underlying target cues were never available to the listener as such, being either masked (level 3) or "replaced" by a da-percept inducing bump (level 1) or a ga-percept inducing bump (level 5). In this respect, the experiment was more similar to a cue-manipulation study than to a SNR-based regression study. Indeed, in the former, the target cues are artificially modified (Clayards, 2018; Hazan & Rosen, 1991; Liberman et al., 1954; Pittman & Stelmachowicz, 2000) while in the latter, the continuum goes from "target cues fully available" to "target cues fully unavailable" (Calandruccio & Doherty, 2007, 2008; Gilbert et al., 2002).

Table 2. Frequency Location of the Bumps Used in the Parametric Bump Noise Experiment.

Cue name	Formant onsets	Center frequency of the bumps	Bias toward
HF cue	F2/F3	2641 Hz	"da"
		2125 Hz	"ga"
		1975 Hz	"ga"
		1500 Hz	"da"
LF cue	F1	578 Hz	"da"
		390 Hz	"ga"

Note. HF = high-frequency region; LF = low-frequency region.

Each condition was repeated 10 times for each subject in a random order, yielding 1,000 trials, which were divided into five blocks of 200 trials separated with pauses to avoid mental fatigue. The total duration of the experiment was approximately 2.5 hr.

Analysis

We modeled the relationship between participants' responses and the acoustic content of the stimuli, on a trial-by-trial basis, using a GLM. The model includes an

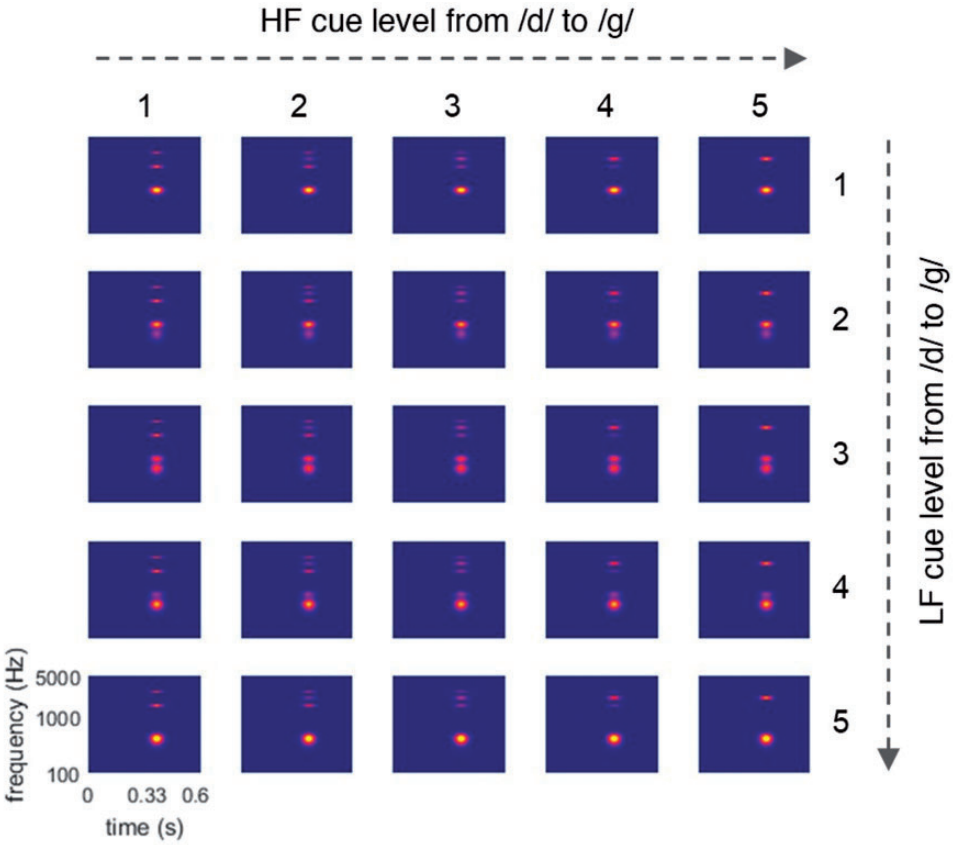

Figure 4. Stimulus design of bump variation along the two-dimensional continuum. Each bump condition corresponds to one of the 5 × 5 ideal time–frequency profile (i.e., to a particular level of the 6 bumps). Arrows indicate the two dimensions along which the bump noise was varied, with the corresponding level of the bump.

effect of HF cue and LF cue (β_{HF} and β_{LF}, respectively), an interaction effect between them ($\beta_{HF\times LF}$), and a four-level factor corresponding to the target actually presented (α_k with $k \in 1,4$).

Let r_j denote the response of one participant to trial j (1 for "da," 0 for "ga"). The probability of a "da" response is given by:

$$\mathrm{P}(r_j = 1) = \Phi(\beta_{HF} \cdot lvlHF_j + \beta_{LF} \cdot lvlLF_j + \beta_{HF\times LF} \cdot lvlHF_j \cdot lvlLF_j + \alpha_{t_j}) \quad (2)$$

with $lvlHF_j$ and $lvlLF_j$ the levels of the bumps superimposed with HF and LF cues, respectively, and t_j the number of the target presented. Variables *lvlHF* and *lvlLF* are centered and normalized. Φ denotes the logit function. In this model, the general bias of a listener in favor of "da" or "ga" materializes as $\beta_{bias} = mean(\underline{\alpha})$.

By construction of the bump continuum, β_{HF} and β_{LF} cannot take negative values. Therefore, they were assigned with log-normal priors. More precisely, the parameters to be estimated are $\log\beta_{HF} = \log(\beta_{HF})$ and $\log\beta_{LF} = \log(\beta_{LF})$. These new parameters $\log\beta_{HF}$ and $\log\beta_{LF}$ were associated with Gaussian priors, as well as all other parameters in the model.

Each participant in the experiment was described by an individual set of parameters $\{\beta_{HF}, \beta_{LF}, \beta_{HF\times LF}, \underline{\alpha}\}$. The dependencies between data from different listeners were accounted for by using hierarchical modelling. As random-effects models in frequentist terms, hierarchical models not only allow the estimation of individual parameters but also take into account their similarities. More specifically, we assumed here that each individual coefficient is drawn from a common distribution corresponding to his group, and that the three group distributions are in turn drawn from a single general distribution. Estimating group and population parameters and using them as priors in a three-level hierarchical model allows pooling the information across individuals, rather than treating them as independent measurements, and improves accuracy.

The computed values of β_{HF} and β_{LF} are estimates of the "true" weights used in the decision process, B_{HF} and B_{LF}, but they additionally incorporate the effect of internal noise (i.e., the stochastic part of the decision process), which acts as a factor on all weights ($\beta_{HF} = \alpha \cdot B_{HF}$ and $\beta_{LF} = \alpha \cdot B_{LF}$) (Kurki et al., 2014; Murray, 2011; Richards & Zhu, 1994). Since in this study we were interested only in the *relative* (rather than *absolute*) decision weights, in the analysis, we focused exclusively on the weight ratios, $\beta_{HF}/\beta_{LF} = B_{HF}/B_{LF}$. Note that the internal noise factors out in the division, so that the preceding equality holds regardless of the magnitude of the internal noise. Because weight ratios do not, in general, have a Gaussian distribution, we actually used log-transformed ratio, $\log(\beta_{HF}/\beta_{LF})$.

The distribution of individual SNRs, scores, and biases were modeled with three separate three-level hierarchical Bayesian models. A simple regression with Gaussian(0,1) priors on the mean values was used for the SNRs while a logistic regression with Gaussian (0,1) priors on the log-odds was used for the scores and biases.

All Bayesian analyses were conducted using JAGS (Plummer, 2003). Seven chains were run independently with 2,000 burn-in samples (estimates based on 8,000 samples in each chain) and were checked visually for convergence. Throughout this article, Bayesian estimates will be reported along with their 95% credible intervals, providing an assessment of the reliability of the estimate.

Results

Behavioral Results

All participants were included in the final analysis. On average, the experiment lasted approximately 3 hr per HI participants and 2.5 hr per NH participants.

Despite partial restoration of audibility through frequency-dependent amplification (NAL-R), HI participants performed more poorly than NH participants in the phoneme-categorization task. HI participants usually needed a higher SNR than NH participants to perform the task at similar performance level (see Figure 5(a)). Individual SNRs spanned values between −6.5 and −13 dB for the NH group, between −3.0 and −10.6 dB for the HF-loss group, and between −1.7 and −10.6 dB for the flat-loss group. According to the Bayesian model on the individual SNR values, the probability that the SNRs from the NH and HI groups come from the same distribution was lower than 0.05 (as also suggested by the disjoint credibility intervals in Figure 5(a)).

In the main experiment, the average percentage of correct responses was 58.8%, whereas the correct recognition scores in white noise targeted by the initial staircase was 70.7% (see Table 1). Therefore, the effect of the addition of bumps in the spectrotemporal regions corresponding to the acoustic cues can be estimated at approximately 12 percentage points. The NH group obtained an average of 56.2% correct, against 60.0% for the HF-loss group and 61.0% for the flat-loss group (see Figure 5(b)). There was a large variability in the individual results (from approximately 50% correct up to 72.3% correct). Note, however, that scores near chance level (50%) may not imply that a participant is responding at random, but only that his or her decision is not driven by the target actually presented. His or her responses may depend on other factors. In particular, it could be influenced by the bump

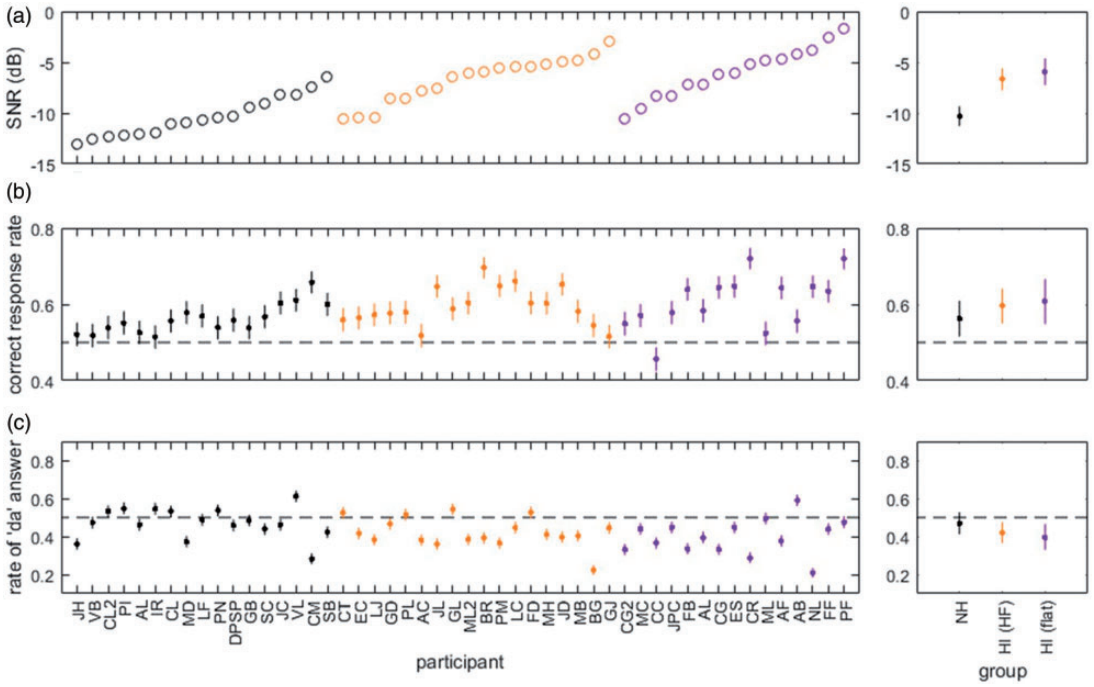

Figure 5. Behavioral results for each participant and each group. (a) SNR thresholds measured in the first phase of the experiment. (b) Performances in the second phase of the experiment (correct response rate). Chance level (50%) is represented with a dashed line. (c) General bias (rate of "da" answers). Fifty percent (dashed line) corresponds to an unbiased behavior. Participants are ordered according to group (black: NH; orange: HI with HF loss; indigo: HI with flat loss) and per SNR. Circles represent the individual SNR thresholds and dots represent the outcome of hierarchical Bayesian models, with 95% credible intervals.

noises, with "da"-like bumps inducing more "da"-like responses than "ga"-like bumps, and vice-versa. According to the Bayesian model on individual scores, there was no strong (<5% chance) group difference between the three groups.

Surprisingly, there was a large interindividual variability in the response bias, as depicted in Figure 5(c), with some participants strongly biased toward response "ga" while others were biased toward "da." Again, no strong difference was found at the group level when modeling the data with a three-level hierarchical logistic regression. The same variability was found at the target level. This is, for example, the case for the "Alda" target, which was mostly perceived as "da" by 43 participants, and mostly as "ga" by six participants.

Cue Sensitivity Analysis

A GLM was fitted on the data from the main experiment to link the amplitude of the bumps ("HF cue level" and "LF cue level") to the percentage of da responses, as described in the "Methods" section. The model included two weights corresponding to the effects of HF and LF cues (β_{HF} and β_{LF}), an interaction between the two ($\beta_{HF\times LF}$), and a bias factor with one level for each of the four possible targets ($\underline{\alpha}$).

Overall, the GLM was quite good at predicting the individual participants' responses. Figure 6 plots the data for each listener (dots) and the model's predictions (lines), averaged across the four targets. In each panel, showing the proportion of "da" answer as a function of LF cue level, with HF cue level as a parameter, the influence of HF cue is reflected by the spacing between lines and the influence of LF cue as the slope of the lines. There was a very good match between the data and the model overall, with less than 4.2% mean absolute error across all listeners and conditions.

Figure 7 represents the values of parameters β_{HF}, β_{LF} and $\beta_{HF\times LF}$ for each level and each group (Panels A, B, and C, respectively). As expected, the weight associated to the HF cue was clearly higher than the one associated to the LF cue (by a ratio of approximately 7.4, across all groups and participants). This is consistent with the idea that the F2–F3 onset is used as a primary cue for this task, whereas the F1 onset plays only a secondary role in the decision. Furthermore, the Bayesian analysis suggests that there is no strong interaction effect between the two cues (the credible intervals for $\beta_{HF\times LF}$ overlap with zero for 43/50 participants), consistent with a model in which the two sources of information are combined linearly for most of the listeners. This is confirmed by the deviance information criterion (DIC), which is a

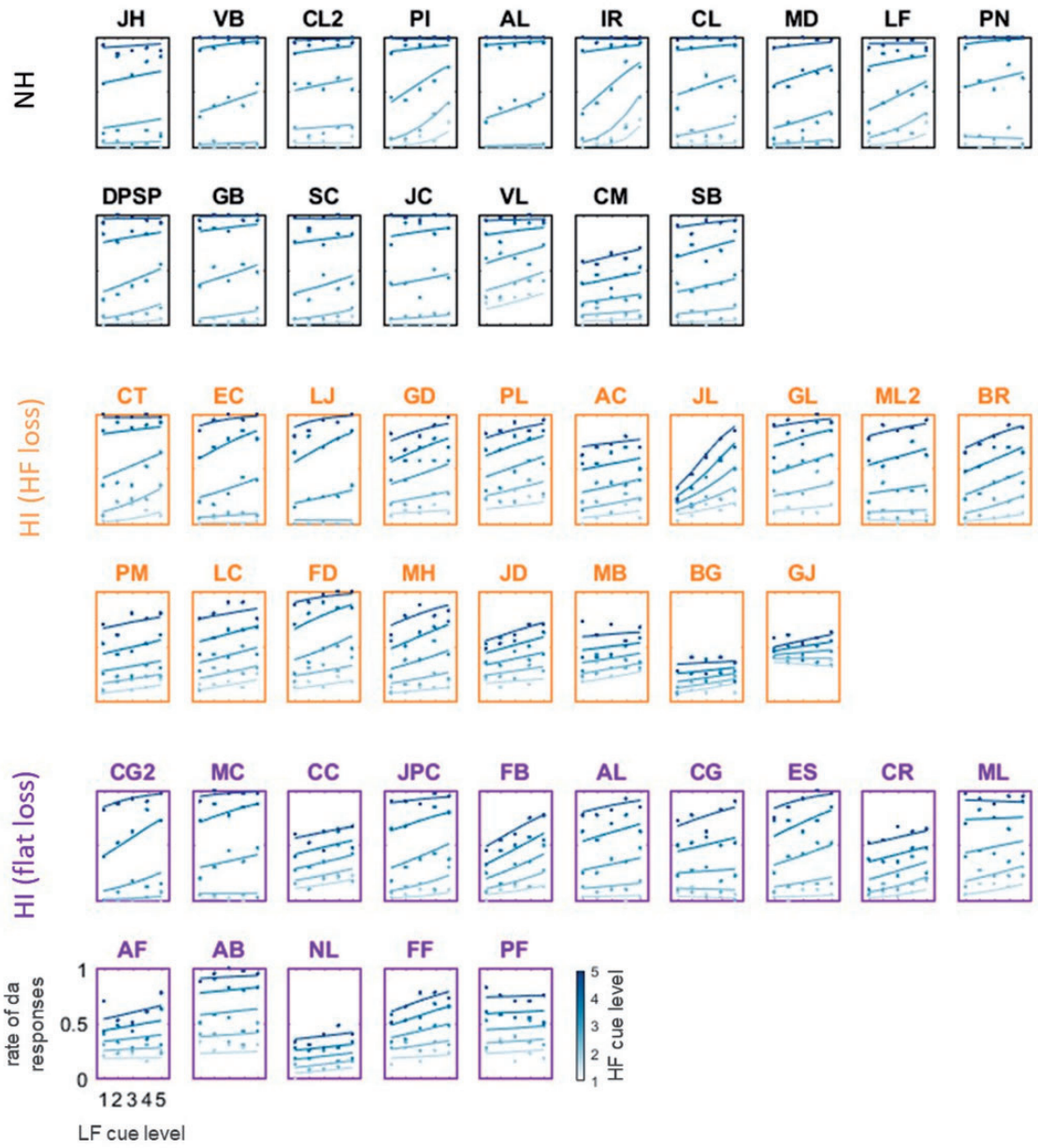

Figure 6. Data measured for all participants (dots) and predictions of the model (lines), averaged across targets. The proportion of "da" answer is plotted as a function of LF cue level, with HF cue level as a parameter (shade of blue).

measure of accuracy for a Bayesian model accounting for overfitting, and is therefore useful for comparing models with different numbers of parameters (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). The DIC was $1.272{\cdot}10^4$ for the model described earlier and $1.275{\cdot}10^4$ for the same model without interaction effect. As suggested by the very small difference between the two, including an interaction parameter in the model does not add much to its predictions. On the contrary, the effect of LF cue, although limited, is necessary from this point of view, as removing the β_{LF} parameter results in a relatively large increase in DIC (DIC $= 1.345{\cdot}10^4$).

The main goal of this study was to compare the *relative* weightings of each cue for the three groups of listeners. For this purpose, we computed the log ratio between the weights associated to HF cue or LF cue (see Figure 7(d)). At the group level, a difference is observed between the log ratios of NH listeners (mean log ratio $= 1.0$) and HI (HF) listeners (mean log ratio $= 0.76$)—The probability that the two groups have different log ratios is above 95%, according to the model. The log ratio for the HI (flat) group (mean log ratio $= 0.87$) has an intermediate value between those for the NH and HI (HF) groups.

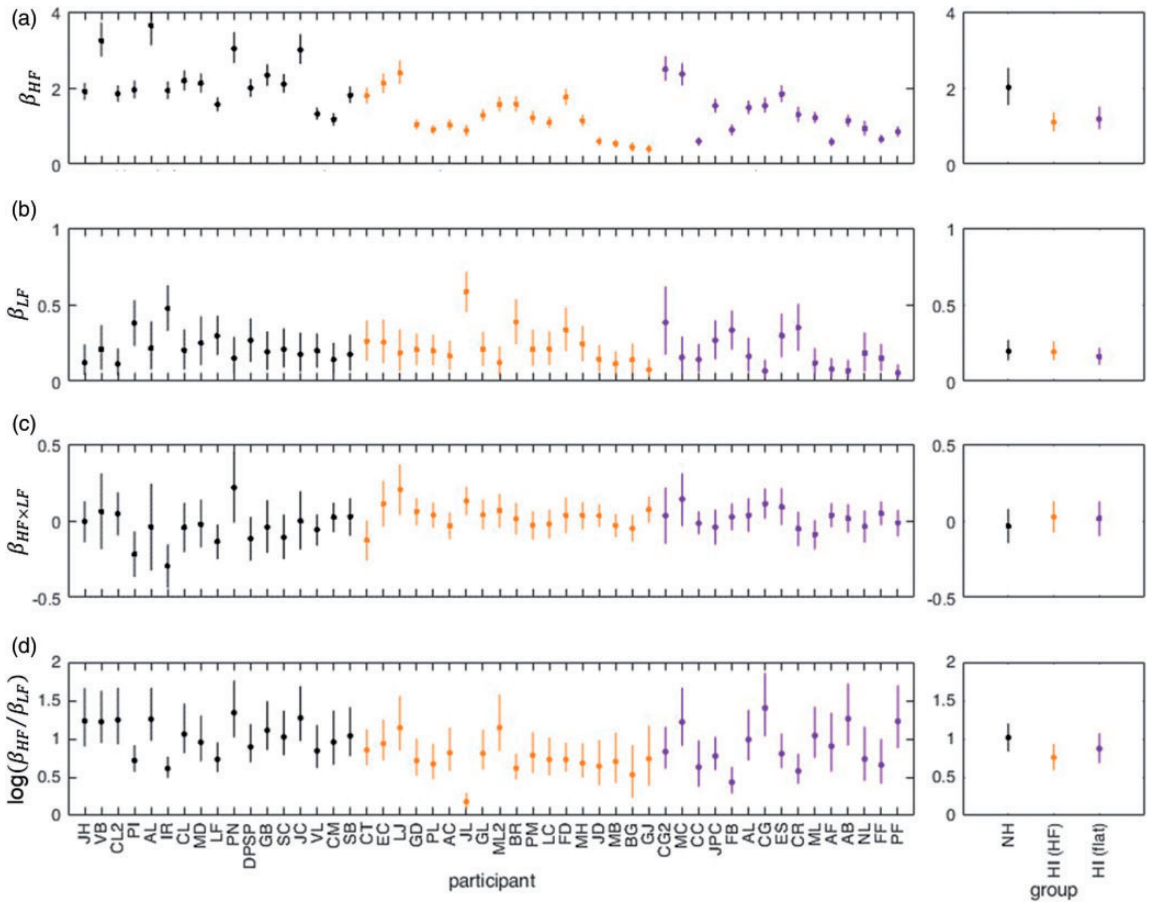

Figure 7. Estimated values for parameters β_{HF} (Panel A), β_{LF} (Panel B), and $\beta_{HF\times LF}$ (Panel C) at the individual and group levels, with 95% credible intervals. The lower panel shows the values of log ratio $\log(\beta_{HF}/\beta_{LF})$. Note the different *y*-axis scales used in the three panels.

Discussion

This study aimed at building upon previous research, which showed that the F2/F3 and F1 onsets are used as acoustic cues for stop-consonant perception in noise (Delattre et al., 1955; Mann, 1980; Varnet, Knoblauch, et al., 2015). Specifically, this study sought to estimate the relative weights of these two cues, thus determining listening strategies in normal- and impaired-hearing listeners for consonant-in-noise comprehensions. To do so, we set up two experiments, both based on recordings of natural speech signals masked by the addition of a "bump noise." The bumps were narrowband bursts of noise placed at the onset of the consonant. In the first (pilot) experiment, the bump noise was used as a mean to determine the frequency locations of the two acoustic cues. In the main experiment, it was designed to alter the perception of consonant (changing "d" into "g" or vice versa) by manipulating specifically these cues, therefore allowing us to measure their respective weights.

As clearly revealed by the differences in overall performance across the two experiments, the bumps had a strong deleterious impact on intelligibility. Each experiment was composed of two phases, the first one (adaptive staircase) using white noise and the other using bump noise at the SNR threshold determined in the first phase. Therefore, the effect of masking on intelligibility due to the addition of the bumps can be estimated by computing the differences in scores across in the two phases. In presence of a bump noise, the percentage of correct answers decreased by approximately 10 to 12 percentage points on average, across groups.

However, the purpose of using bump noise was not only to impair intelligibility, but rather to manipulate the listener's phonetic percept. As revealed by the bump distributions in Figure 2, depending on its spectral content, the bump noise biased the listener's percept toward "da" or "ga." On the whole, bumps placed at time–frequency positions corresponding to an acoustic cue tended to enhance the percept normally induced by this cue. The two frequencies where the presence of a bump had the strongest impact on the listener's responses are 1500 Hz (for the "da" bump) and 1975 Hz (for the "ga" bump). As may be seen on the spectrograms of the targets displayed in Figure 1, these frequencies match those of the F2 onsets in the syllables

/da/ and /ga/, respectively. Therefore, a bump noise containing more energy around 1500 Hz probably made the onset of the second formant be perceived as lower than it actually was, and the participant was more likely to answer "da." This is consistent with the idea that the frequency of F2 at onset is an important cue for stop-consonant categorization (Delattre et al., 1955; Liberman et al., 1954). Following the same reasoning, the pilot experiment revealed that two broad regions play a critically important role in influencing the listener's decision: a high-frequency region (HF cue, between 1400 and 2700 Hz) corresponding to the F2 and F3 onsets and a weaker low-frequency region (LF cue, between 350 and 600 Hz) corresponding to the F1 onset (shaded regions in Figure 2). These observations are consistent with the results obtained by Varnet et al. on the same stimuli, using a different psychophysical revcorr method (Varnet, Knoblauch, et al., 2015; Varnet, Meunier, Trollé, et al., 2016; Varnet, Wang, et al., 2015). As highlighted in the introduction, this listening strategy may be specific to speech-in-noise comprehension. In quiet, however, additional cues, such as burst cues, may be used (Kapoor & Allen, 2012).

The implication of the two aforementioned cues was further confirmed in the main experiment. Here, the positions of the bumps were fixed on the six most critical frequencies listed in Table 2, but their amplitude were varied, so as to create a 5×5 continuum (full factorial design with five factorial levels on each of the two cues, see Figure 4). This particular type of noise appears to have a dramatic effect on perception. As can be seen on Figure 6, for most of the NH participants, varying the amplitude of the bumps on HF cue from Level 1 (light blue dots) to Level 5 (dark blue dots) increased the percentage of "da" responses from almost 0% to near 100%. The LF cue level factor appears to have a weaker effect on perception, as indicated by the shallower psychometric functions. When HF cue is ambiguous (level = 3), the variation of the bump amplitude on LF cue induces a 12.5% change in the percentage of "da" responses from NH participants, on average. This may be related to Delattre's observation that "when the straight second formant is about midway between the g locus (at 3000 cps) and the d locus (at 1800 cps), raising or lowering the level of the first formant tends to push the sounds toward d or g" (Delattre et al., 1955, p. 3).

To evaluate quantitatively the relative influence of the low- and high-frequency cues, we fitted a GLM to the participants' data (see "Methods" section). The model included an effect of HF cue and LF cue (β_{HF} and β_{LF}, respectively), an interaction effect between them ($\beta_{HF\times LF}$), and a four-level factor corresponding to the target actually presented (α_k with $k \in 1,4$). The fit of the model was very good (with a mean absolute error of less than 4.2 percentage points across all listeners and

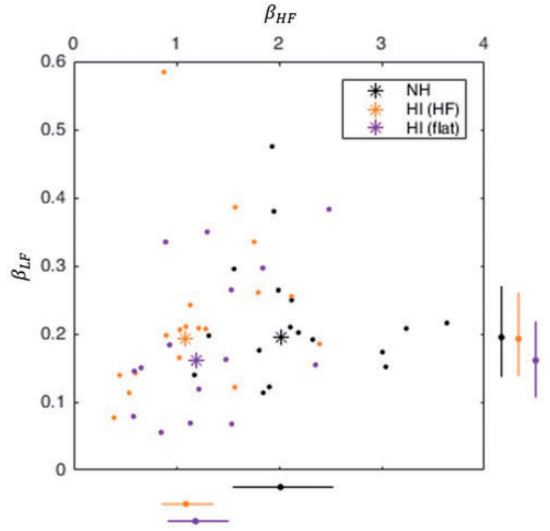

Figure 8. Scatterplot of β_{HF} as a function of β_{LF}, at the individual (dots) and group (stars) levels.

conditions), supporting the idea that this model, although simple, already provides a good account of the data. Consistent with previous observations on the secondary role of the F1 cue in the da/ga categorization in noise (Varnet, Knoblauch, et al., 2015), the weight of the HF cue in this model was always stronger (by a factor of about 7) than the weight of the LF cue. The scatterplot of β_{LF} as a function of β_{HF}, presented in Figure 8, gives a graphical representation of the ratio between the two cues as well as the variability between participants.

For most of the participants, the interaction effect, if any, was small, suggesting a linear combination of two cues (prior to nonlinear transformation of the decision variable by the logistic link function). Various authors have assumed that secondary cues were used only when primary cues were unreliable or removed (Delattre et al., 1955; Li et al., 2010; Serniclaes & Arrouas, 1995). If this assumption of a binary process was true, one would expect the interaction term in our model to be significant. On the contrary, the small values of $\beta_{HF\times LF}$ observed in all participants point to a *constant* contribution of the secondary cue to the internal decision variable, whatever the primary cue level. According to this view, the observation by Delattre that secondary cues are used, or not, conditionally on the value of the primary cue, may be an artifact due to the percentage representation, which introduces floor and ceiling effects when the primary cue is at one end of the continuum.

Although the same listening strategy (differential weighting of HF and LF cues with no or very little interaction between them) is shared by all 17 participants in the NH group, there is still a large heterogeneity in the

exact values of the estimated weights in the individual models (Figures 7 and 8). This may indicate idiosyncratic differences in the processing of the cues by listeners (i.e., in the actual weights used by NH listeners). Alternatively, this variability may arise from a different, later-occurring factor, such as attentional effects. For example, all other things being equal, the estimated weights of the two cues will be smaller overall if a participant is less focused on the task, yielding more variable responses (i.e., more internal noise). Within the standard signal detection theory framework, these effects are modeled as a source of noise added to the internal decision variable (Green, 1964; Neri, 2013). Taking the log ratio of the two cues allowed us to factor out the effect of internal noise.

The main objective of this article was to compare the log ratio values of the NH group with those obtained by listeners with hearing loss. Two types of audiometric configurations were considered: high-frequency loss (HI with HF-loss group) and flat or gradually sloping loss (HI with flat-loss group). For these two groups, stimuli were presented at the same overall level as for NH listeners (70 dB SPL), but the sounds were processed through a simulated hearing aid adjusted to individual hearing loss profile according to the NAL-R formula (Byrne & Dillon, 1986; Palmer & Lindley, 2002). We reasoned that if the three groups used different listening strategies, this should be reflected in an assessable difference in log ratios at the group level. As a matter of fact, Figure 7(d) reveals that HI participants with a HF loss, as a group, relied less on the high-frequency cue than on the low-frequency one, compared with NH listeners, even though stimuli were amplified in order to partially compensate for audibility.

This result contrasts with those of previous studies using correlational methods on individuals with hearing impairment. In speech (Calandruccio & Doherty, 2008; Gilbert et al., 2002) as in nonspeech (Doherty & Lutfi, 1996, 1999) tasks, participants with HF loss appear to weight high-frequency information more heavily, even when audibility is restored through a simulated hearing aid. Authors have interpreted this as an attempt of HI listeners to compensate for the degraded sensory information by using a different listening strategy. However, in the aforementioned studies, the two groups were matched in terms of performances but not in overall presentation level. HI listeners were presented with more energetic (and possibly louder) stimuli on average in order to partially compensate for their hearing loss and to reach the desired performance level. Therefore, the increased weighting of the HF regions could be explained by a difference in presentation levels alone (Calandruccio, Buss, & Doherty, 2016; Jesteadt, Valente, Joshi, Schmid, 2014; Leibold, Tan, & Jesteadt, 2009; Lentz & Leek, 2002). To avoid this potential pitfall, the present study was carried out with NH and HI participants listening at the same overall level of 70 dB SPL. The results suggest a greater reliance on low-frequency information in HI listeners with high-frequency hearing loss.

A possible explanation for this result is that any residual effect of hearing loss after the partial compensation of audibility by frequency-dependent amplification (with NAL-R) was still large enough for acoustic cues falling into the frequency region of the loss to be less reliable or, at least, less relied upon by the listener. Such an explanation would be in agreement with previous works showing that HI listeners are not fully able to make use of available (audible) acoustic cues (Trevino & Allen, 2013; Turner & Brus, 2001; Turner et al., 1992; Turner & Robb, 1987), a phenomenon often cited as evidence for "supra-threshold" deficits (Léger, Moore, & Lorenzi, 2012; Plomp, 1978). From this point of view, the results of this study may shed some light on why HI individuals often have difficulties correctly identifying consonants, even when wearing their hearing aids (Abavisani & Allen, 2017; Scheidiger & Allen, 2013; Scheidiger et al., 2017).

Although the overall presentation levels of the stimuli were equalized across the three groups, another potential confounding factor must be considered in the interpretation of the results. As apparent in Figure 5(a), the SNR at which each group was able to perform the task at 70.7% correct in white noise was markedly different (average 4 dB SNR difference between the NH group and the HI groups). Furthermore, the individual correct response rates in the main experiment are strongly correlated with SNR levels, as revealed by a hierarchical Bayesian logit regression model between percent correct scores and SNR with Gaussian(0,1) priors on model parameters (credible interval above zero for the slope parameters for the three groups). Figure 9(a) plots the individual correct response rates as a function of SNR, as well as the regression curves and the posterior predictive confidence intervals (dotted lines). The relationship between SNR and performance in bump noise is primarily due to the fact that SNR thresholds were measured in white noise. As bump levels were specified relative to the baseline noise level, participants performing at lower SNRs were confronted with bump noise more deleterious to intelligibility than participants at high SNRs. This may not be an issue here, however, as our main interest is on the relative importance of the cues, $\log(\beta_{HF}/\beta_{LF})$, and not on the absolute magnitude of the weights β_{HF} and β_{LF}. Yet another phenomenon, such as an adaptation of the listening strategy to the level of background noise, may have come into play. The influence of SNR on the log ratio was assessed by means of a hierarchical Bayesian regression model taking into account the uncertainty in

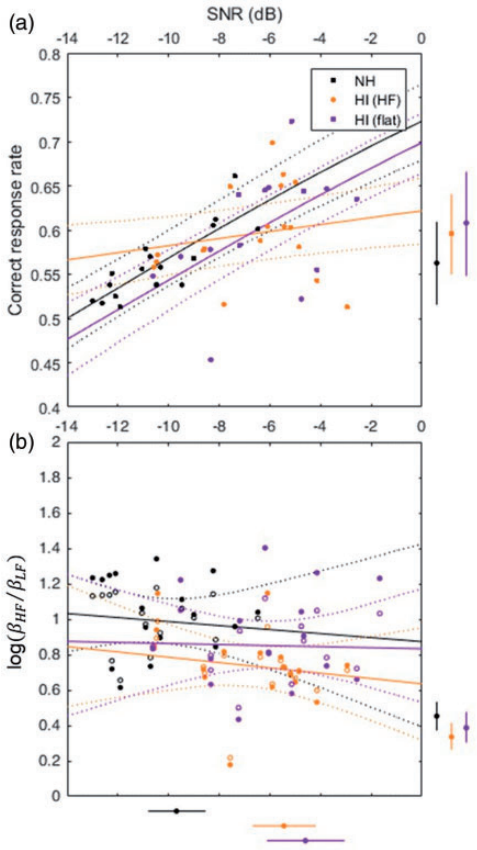

Figure 9. Correlations between correct response rate and SNR (a) and between log ratios and SNR (b), assessed with two hierarchical Bayesian regression models. The dotted lines correspond to the 95% posterior predictive confidence intervals. The open circles in panel b represents the "true" log ratios inferred by the regression model (see text).

the estimation of the log ratio, as described in Matzke et al. (2017). The strength of the correlation measured in this way is always higher than the traditional point-wise Pearson's correlation coefficient because it allows for shrinkage of individual data points toward the regression line. In Figure 9(b), the dots represent the original log ratio estimates from Figure 7(d). whereas the open circles correspond to the "true" values inferred by the regression. Even taking into account this source of uncertainty, the correlation between SNR thresholds and log ratios was very weak (see Figure 9(b)), with the credible intervals for the three groups' slope coefficients intersecting zero. Therefore, we can exclude that individual variations in SNR thresholds explain the observed difference in listening strategies between HI and NH groups.

From a methodological point of view, this study introduces a new type of noise, called "bump noise," that allows the experimenter to control the percept of consonants in noise through the manipulation of acoustic cues. We demonstrated two potential applications of such bump noise. First, the "bump noise ACI" technique, used in the pilot experiment, is a revcorr approach based on the presentation of randomly located bump. As noted in the "Methods" section, this approach is very close conceptually to the "white noise ACI" method described in Varnet et al., (2015). Despite the power and flexibility of the latter, its implementation is limited by the amount of data that can be obtained in a given psychoacoustical task. Using a bump noise instead of a white noise effectively reduces the dimensionality of the problem, and therefore the duration of the experiment, by introducing additional assumptions in the process (here, the fact that the acoustic cues been sought are located on the onset of the second syllable and have a width of at least 1 ERB_N). For example, the white noise ACI experiment on the da/ga categorization task required 10,000 trials per participant (Varnet, Knoblauch, et al., 2015) whereas 1,000 trials per participants were sufficient in the pilot experiment using a bump noise ACI approach on the same task. In this study, we decided to use white noise as basis for the bump noise, in order to stay as close as possible to the original white noise ACI experiments which inspired this work (Varnet, Knoblauch, et al., 2015). However, flat spectral densities are very uncommon in natural sounds. Furthermore, spectral distribution of the masker is likely to affect the listening strategies, as high- and low-frequency cues will be differently reliable (Phatak, Lovitt, & Allen, 2008). Further experiments (e.g., using speech-shaped bump noise) could be carried in the future to quantify the change in cue-weighting strategy depending on the type of noise encountered.

In the main experiment, the bump noise content was varied in a more parametric way to estimate the psychometric functions associated to each acoustic cue. Although previous studies have already explored qualitatively the relationship between primary and secondary cues using continua of modified speech signals (e.g., Delattre et al., 1955; Li et al., 2010; Ohde & Stevens, 1983; Serniclaes & Arrouas, 1995), few have attempted to estimate quantitatively the relative weightings of different cues in a phoneme categorization task (Clayards, 2018; Gilbertson & Lutfi, 2014; Hazan & Rosen, 1991; Pittman & Stelmachowicz, 2000; Pittman et al., 2002). These two approaches provide insights into the individual HI listener's perceptual strategy. As such, the observed variability in cue weighting despite restored audibility points toward the need for more individually tailored speech-in-noise processing in hearing aids.

Data Accessibility Statement

The data supporting the findings of this study are openly available as supplementary materials.

Declaration of Conflicting of Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: C. M. is supported by a for-profit hearing aid manufacturer, Starkey. Other than through this author's contributions to the formulation of the study goals and hypotheses, the design of the experiment and of the statistical model, and the writing of the manuscript, this company had no influence on the work.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Léo Varnet, Chloé Langlet, and Christian Lorenzi are supported by the EUR Frontiers in Cognition grant ANR-17-EURE-0017.

ORCID iD

Léo Varnet https://orcid.org/0000-0002-9702-2649

References

Abavisani, A., & Allen J. B. (2017). Evaluating hearing aid amplification using idiosyncratic consonant errors. *The Journal of the Acoustical Society of America*, *142*(6), 3736. doi:10.1121/1.5016852.

Allen, J. B. (1994). How do humans process and recognize speech? *IEEE Transactions on Speech and Audio Processing*, *2*(4), 567–577. doi:10.1109/89.326615.

American National Standards Institute. (1997). *American National Standard Methods for the Calculation of the Speech Intelligibility Index (ANSI S3.5-1997)*. New York, NY: Author.

Bernstein, J. G. W., Danielsson, H., Hällgren M., Stenfelt, S., Rönnberg, J., & Lunner, T. (2016). Spectrotemporal modulation sensitivity as a predictor of speech-reception performance in noise with hearing aids. *Trends in Hearing*, *20*. doi:10.1177/2331216516670387.

Bilger, R. C., & Wang, M. D. (1976). Consonant confusions in patients with sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research*, *19*(4): 718–748. doi:10.1044/jshr.1904.718.

Brimijoin, W. O., Akeroyd, M. A., Tilbury, E., & Porr, B. (2013). The internal representation of vowel spectra investigated using behavioral response-triggered averaging. *The Journal of the Acoustical Society of America*, *133*(2), EL118–122. doi:10.1121/1.4778264.

Byrne, D., & Dillon, H. (1986). The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing*, *7*(4), 257–265. doi:10.1097/00003446-198608000-00007.

Calandruccio, L., & Doherty, K. A. (2007). Spectral weighting strategies for sentences measured by a correlational method. *The Journal of the Acoustical Society of America*, *121*(6), 3827–3836. doi:10.1121/1.2722211.

Calandruccio, L., & Doherty, K. A. (2008). Spectral weighting strategies for hearing-impaired listeners measured using a correlational method. *The Journal of the Acoustical Society of America*, *123*(4), 2367–2378. doi:10.1121/1.2887857.

Calandruccio, L., Buss, E., & Doherty, K. A. (2016). The effect of presentation level on spectral weights for sentences. *The Journal of the Acoustical Society of America*, *139*(1), 466–471. doi:10.1121/1.4940211.

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *The Journal of the Acoustical Society of America*, *144*(3), EL172–EL177. doi:10.1121/1.5052025.

Delattre, P. (1968). From acoustic cues to distinctive features. *Phonetica*, *18*(4), 198–230. doi:10.1159/000258610.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, *27*(4), 769–773. doi:10.1121/1.1908024.

Doherty, K. A., & Lutfi, R. A. (1996). Spectral weights for overall level discrimination in listeners with sensorineural hearing loss. *The Journal of the Acoustical Society of America*, *99*(2), 1053–1058. doi:10.1121/1.414634.

Doherty, K. A., & Lutfi, R. A. (1999). Level discrimination of single tones in a multitone complex by normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *105*(3), 1831–1840. doi:10.1121/1.426742.

Dorman, M. F., Lindholm, J. M., & Hannley, M. T. (1985). Influence of the first formant on the recognition of voiced stop consonants by hearing-impaired listeners. *Journal of Speech and Hearing Research*, *28*(3), 377–380. doi:10.1044/jshr.2803.377.

Dubno, J. R., Dirks, D. D., & Langhofer, L. R. (1982). Evaluation of hearing-impaired listeners using a Nonsense-syllable Test. II. Syllable recognition and consonant confusion patterns. *Journal of Speech and Hearing Research*, *25*(1), 141–148. doi:10.1044/jshr.2501.141.

Ewert, S. D. (2013). *AFC—A modular framework for running psychoacoustic experiments and computational perception models*. Paper presented at Proceedings of AIA-DAGA March 2013, Merano, Italy

Gilbert, G., Micheyl, C., Berger-Vachon, C., Collet, L. (2002, September). Frequency-importance functions for speech in young and older listeners. Paper presented at Forum Acousticum, Seville, France.

Gilbertson, L., & Lutfi, R. A. (2014). Correlations of decision weights and cognitive function for the masked discrimination of vowels by young and old adults. *Hearing Research*, *317*, 9–14. doi:10.1016/j.heares.2014.09.001.

Glasberg, B. R., & Moore, B. C. (1989). Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scandinavian Audiology. Supplementum*, *32*, 1–25.

Green, D. M. (1964). Consistency of auditory detection judgments. *Psychological Review*, *71*(5), 392–407. doi:10.1037/h0044520.

Hazan, V., & Rosen, S. (1991). Individual variability in the perception of cues to place contrasts in initial stops. *Perception & Psychophysics*, *49*(2), 187–200. doi:10.3758/BF03205038.

Hogan, C. A., & Turner, C. W. (1998). High-frequency audibility: Benefits for hearing-impaired listeners. *The Journal of*

the Acoustical Society of America, *104*(1), 432–441. doi:10.1121/1.423247.

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

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, *16*(4), 305–312. doi:10.1111/j.0956-7976.2005.01532.x.

Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *Journal of the American Academy of Audiology*, *18*(7), 590–603.

Jesteadt, W, Valente, D. L., Joshi, S. N., & Schmid, K. K. (2014). Perceptual weights for loudness judgments of six-tone complexes. *The Journal of the Acoustical Society of America*, *136*(2), 728–735. doi:10.1121/1.4887478.

Kapoor, A., & Allen, J. B. (2012). Perceptual effects of plosive feature modification. *The Journal of the Acoustical Society of America*, *131*(1), 478–491. doi:10.1121/1.3665991.

Kurki, I., & Eckstein, M. P. (2014). Template changes with perceptual learning are driven by feature informativeness. *Journal of Vision*, *14*(11), 6. doi:10.1167/14.11.6.

Kurki, I., Saarinen, J., & Hyvärinen, A. (2014). Investigating shape perception by classification images. *Journal of Vision*, *14*(12), 24. doi:10.1167/14.12.24.

Léger, A. C., Moore, B. C. J., & Lorenzi, C. (2012). Abnormal speech processing in frequency regions where absolute thresholds are normal for listeners with high-frequency hearing loss. *Hearing Research*, *294*(1–2), 95–103. doi:10.1016/j.heares.2012.10.002.

Leibold, L. J., Tan, H., & Jesteadt, W. (2009). Spectral weights for sample discrimination as a function of overall level. *The Journal of the Acoustical Society of America*, *125*(1), 339–346. doi:10.1121/1.3033741.

Lentz, J. J., & Leek, M. R. (2002). Decision strategies of hearing-impaired listeners in spectral shape discrimination. *The Journal of the Acoustical Society of America*, *111*(3), 1389–1398.

Lesica, N. A. (2018). Why do hearing aids fail to restore normal auditory perception? *Trends in Neurosciences*, *41*(4), 174–185. doi:10.1016/j.tins.2018.01.008.

Li, F., & Allen, J. B. (2011). Manipulation of consonants in natural speech. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *19*(3), 496–504. doi:10.1109/TASL.2010.2050731.

Li, F., Menon, A., & Allen, J. B. (2010). A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *The Journal of the Acoustical Society of America*, *127*(4), 2599–2610. doi:10.1121/1.3295689.

Li, R. W., Klein, S. A., & Levi, D. M. (2006). The receptive field and internal noise for position acuity change with feature separation. *Journal of Vision*, *6*(4), 311–321. doi:10.1167/6.4.2.

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

Mackersie, C. L. (2007). Temporal intra-speech masking of plosive bursts: Effects of hearing loss and frequency shaping. *Journal of Speech, Language, and Hearing Research*, *50*(3), 554–563. doi:10.1044/1092-4388(2007/038).

Mandel, M. I., Yoho, S. E., & Healy, E. W. (2016). Measuring time-frequency importance functions of speech with bubble noise. *The Journal of the Acoustical Society of America*, *140*(4), 2542. doi:10.1121/1.4964102.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, *28*(5), 407–412. doi:10.3758/bf03204884.

Matzke. D., Ly, A., Selker, R., Weeda, W. D., Scheibehenne, B., Lee, M. D., & Wagenmakers, E. J. (2017). Bayesian inference for correlations in the presence of measurement error and estimation uncertainty. *Collabra: Psychology*, *3*(1), 25. doi:10.1525/collabra.78.

Moore, B. C. J. (2005). Basic psychophysics of human spectral processing. *International Review of Neurobiology*, *70*, 49–86. doi:10.1016/S0074-7742(05)70002-7.

Moore, B. C. J., & Vinay, S. N. (2009). Enhanced discrimination of low-frequency sounds for subjects with high-frequency dead regions. *Brain: A Journal of Neurology*, *132*(Pt 2), 524–536. doi:10.1093/brain/awn308.

Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, *11*(5), 2. doi:10.1167/11.5.2.

Neri, P. (2013). The statistical distribution of noisy transmission in human sensors. *Journal of Neural Engineering*, *10*(1), 016014. doi:10.1088/1741-2560/10/1/016014.

Ohde, R. N., & Stevens, K. N. (1983). Effect of burst amplitude on the perception of stop consonant place of articulation. *The Journal of the Acoustical Society of America 74*(3): 706–714. doi:10.1121/1.389856.

Owens, E. (1978). Consonant errors and remediation in sensorineural hearing loss. *The Journal of Speech and Hearing Disorders 43*(3): 331–347. doi:10.1044/jshd.4303.331.

Palmer, C. V., & Lindley, G. A. (2002). Overview and rationale for prescriptive formulas for linear and nonlinear hearing aids. In M. Valente (Ed.), *Strategies for selecting and verifying hearing aid fittings* (2nd ed., pp. 1–22). New York, NY: Thieme Medical Publisher.

Phatak, S. A., & Allen, J. B. (2007). Consonant and vowel confusions in speech-weighted noise. *The Journal of the Acoustical Society of America*, *121*(4), 2312–2326. doi:10.1121/1.2642397.

Phatak, S. A., Lovitt, A., & Allen, J. B. (2008). Consonant confusions in white noise. *The Journal of the Acoustical Society of America*, *124*(2), 1220–1233. doi:10.1121/1.2913251.

Phatak, S. A., Yoon, Y., Gooler, D. M., & Allen, J. B. (2009). Consonant recognition loss in hearing impaired listeners. *The Journal of the Acoustical Society of America*, *126*(5), 2683–2694. doi:10.1121/1.3238257.

Pittman, A. L., & Stelmachowicz, P. G. (2000). Perception of voiceless fricatives by normal-hearing and hearing-impaired children and adults. *Journal of Speech, Language, and Hearing Research*, *43*(6), 1389–1401.

Pittman, A. L., Stelmachowicz, P. G., Lewis, D. E., & Hoover, B. M. (2002). Influence of hearing loss on the perceptual strategies of children and adults. *Journal of Speech, Language, and Hearing Research*, *45*(6), 1276–1284. doi:10.1044/1092-4388(2002/102).

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *The Journal of the*

Acoustical Society of America, *63*(2), 533–549. doi:10.1121/1.381753.

Plummer, M. (2003, March). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Paper presented at the 3rd international workshop on distributed statistical computing, Vienna, Austria.

Régnier, M. S., & Allen, J. B. (2008). A method to identify noise-robust perceptual features: Application for consonant/t/. *The Journal of the Acoustical Society of America*, *123*(5), 2801–2814. doi:10.1121/1.2897915.

Richards, V. M., & Zhu, S. (1994). Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients. *The Journal of the Acoustical Society of America*, *95*(1), 423–434. doi:10.1121/1.408336.

Scheidiger, C., & Allen, J. B. (2013). *Effects of NALR on consonant-vowel perception*, Paper presented at the International Symposium on Auditory and Audiological Research, August 2013, Nyborg, Denmark.

Scheidiger, C., Allen, J. B., & Dau, T. (2017). Assessing the efficacy of hearing-aid amplification using a phoneme test. *The Journal of the Acoustical Society of America*, *141*(3), 1739. doi:10.1121/1.4976066.

Seldran, F., Gallego, S., Micheyl, C., Veuillet, E., Truy, E., & Thai-Van, H. (2011). Relationship between age of hearing-loss onset, hearing-loss duration, and speech recognition in individuals with severe-to-profound high-frequency hearing loss. *Journal of the Association for Research in Otolaryngology*, *12*(4), 519–534. doi:10.1007/s10162-011-0261-8.

Serniclaes, W., & Arrouas, Y. (1995). Perception des traits phonétiques dans le bruit (perception of phonetic features in noise). *Verbum* (2): 131–144.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303–304. doi:10.1126/science.270.5234.303.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639. doi:10.1111/1467-9868.00353.

Summers, V., & Leek, M. R. (1997). Intraspeech spread of masking in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *101*(5 Pt 1), 2866–2876. doi: 10.1121/1.419303.

Trevino, A., & Allen, J. B. (2013). Within-consonant perceptual differences in the hearing impaired ear. *The Journal of the Acoustical Society of America*, *134*(1), 607–617. doi:10.1121/1.4807474.

Turner, C. W., & Brus, S. L. (2001). Providing low- and mid-frequency speech information to listeners with sensorineural hearing loss. *The Journal of the Acoustical Society of America*, *109*(6), 2999–3006. doi:10.1121/1.1371757.

Turner, C. W., Fabry, D. A., Barrett, S., & Horwitz, A. R. (1992). Detection and recognition of stop consonants by normal-hearing and hearing-impaired listeners. *Journal of Speech and Hearing Research*, *35*(4), 942–949. doi:10.1044/jshr.3504.942.

Turner, C. W., & Robb, M. P. (1987). Audibility and recognition of stop consonants in normal and hearing-impaired subjects. *The Journal of the Acoustical Society of America*, *81*(5), 1566–1573. doi:10.1121/1.394509.

Varnet, L., Knoblauch, K., Meunier, F., & Hoen, M. (2013). Using auditory classification images for the identification of fine acoustic cues used in speech perception. *Frontiers in Human Neuroscience*, *7*, 865. doi:10.3389/fnhum.2013.00865.

Varnet, L., Knoblauch, K., Serniclaes, W., Meunier, F., & Hoen, M. (2015). A psychophysical imaging method evidencing auditory cue extraction during speech perception: A group analysis of auditory classification images. *PLoS One*, *10*(3), e0118009. doi:10.1371/journal.pone.0118009.

Varnet, L., Wang, T., Peter, C., Meunier, F., & Hoen, M. (2015). How musical expertise shapes speech perception: Evidence from auditory classification images. *Scientific Reports*, *5*, 14489. doi:10.1038/srep14489.

Varnet, L., Meunier, F., Trollé, G., & Hoen, M. (2016). Direct viewing of dyslexics' compensatory strategies in speech in noise using auditory classification images. *PLoS One*, *11*(4), e0153781. doi:10.1371/journal.pone.0153781.

Varnet, L., Meunier, F., & Hoen, M. (2016, September). *Speech reductions cause a de-weighting of secondary acoustic cues*. Paper presented at Interspeech, San Francisco, CA.

Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2010) Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(4), 1005–1015. doi:10.1037/a0018391.