



**HAL**  
open science

# Effect of frequency mismatch and band partitioning on vocal tract length perception in vocoder simulations of cochlear implant processing

Nawal El Boghdady, Deniz Başkent, Etienne Gaudrain

► **To cite this version:**

Nawal El Boghdady, Deniz Başkent, Etienne Gaudrain. Effect of frequency mismatch and band partitioning on vocal tract length perception in vocoder simulations of cochlear implant processing. *Journal of the Acoustical Society of America*, 2018, 143 (6), pp.3505-3519. 10.1121/1.5041261 . hal-02385569

**HAL Id: hal-02385569**

**<https://hal.science/hal-02385569>**

Submitted on 28 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Effect of frequency mismatch and band partitioning on vocal tract length perception in vocoder simulations of cochlear implant processing

Nawal El Boghdady, Deniz Başkent, and Etienne Gaudrain

Citation: [The Journal of the Acoustical Society of America](#) **143**, 3505 (2018); doi: 10.1121/1.5041261

View online: <https://doi.org/10.1121/1.5041261>

View Table of Contents: <http://asa.scitation.org/toc/jas/143/6>

Published by the [Acoustical Society of America](#)

---

## Articles you may be interested in

[Musician effect on perception of spectro-temporally degraded speech, vocal emotion, and music in young adolescents](#)

[The Journal of the Acoustical Society of America](#) **143**, EL311 (2018); 10.1121/1.5034489

[Explaining intelligibility in speech-modulated maskers using acoustic glimpse analysis](#)

[The Journal of the Acoustical Society of America](#) **143**, EL449 (2018); 10.1121/1.5041466

[Human ossicular-joint flexibility transforms the peak amplitude and width of impulsive acoustic stimuli](#)

[The Journal of the Acoustical Society of America](#) **143**, 3418 (2018); 10.1121/1.5039845

[Temporal factors in cochlea-scaled entropy and intensity-based intelligibility predictions](#)

[The Journal of the Acoustical Society of America](#) **143**, EL443 (2018); 10.1121/1.5041468

[Development of speech rhythm in first language: The role of syllable intensity variability](#)

[The Journal of the Acoustical Society of America](#) **143**, EL463 (2018); 10.1121/1.5042083

[Learning to extract a large inter-aural level difference in lag clicks](#)

[The Journal of the Acoustical Society of America](#) **143**, EL456 (2018); 10.1121/1.5041467

---

# Effect of frequency mismatch and band partitioning on vocal tract length perception in vocoder simulations of cochlear implant processing<sup>a)</sup>

Nawal El Boghdady,<sup>b)</sup> Deniz Başkent,<sup>c)</sup> and Etienne Gaudrain<sup>d)</sup>

University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology/Head and Neck Surgery, Groningen, The Netherlands

(Received 17 January 2018; revised 20 April 2018; accepted 22 May 2018; published online 12 June 2018)

The vocal tract length (VTL) of a speaker is an important voice cue that aids speech intelligibility in multi-talker situations. However, cochlear implant (CI) users demonstrate poor VTL sensitivity. This may be partially caused by the mismatch between frequencies received by the implant and those corresponding to places of stimulation along the cochlea. This mismatch can distort formant spacing, where VTL cues are encoded. In this study, the effects of frequency mismatch and band partitioning on VTL sensitivity were investigated in normal hearing listeners with vocoder simulations of CI processing. The hypotheses were that VTL sensitivity may be reduced by increased frequency mismatch and insufficient spectral resolution in how the frequency range is partitioned, specifically where formants lie. Moreover, optimal band partitioning might mitigate the detrimental effects of frequency mismatch on VTL sensitivity. Results showed that VTL sensitivity decreased with increased frequency mismatch and reduced spectral resolution near the low frequencies of the band partitioning map. Band partitioning was independent of mismatch, indicating that if a given partitioning is suboptimal, a better partitioning might improve VTL sensitivity despite the degree of mismatch. These findings suggest that customizing the frequency partitioning map may enhance VTL perception in individual CI users. © 2018 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/10.1121/1.5041261>

[JFL]

Pages: 3505–3519

## I. INTRODUCTION

In individuals with profound sensorineural hearing loss, functional hearing can be restored with the help of a multi-channel cochlear implant (CI): a neural prosthetic device that electrically stimulates the auditory nerve fibres. Currently, while speech perception in quiet is usually good for most CI users (Blamey *et al.*, 2012; Dowell *et al.*, 1986; Tyler *et al.*, 1988), a major challenge lies in understanding speech in the presence of another competing talker (e.g., Pyschny *et al.*, 2011; Stickney *et al.*, 2004). In contrast, normal hearing (NH) listeners can understand speech relatively well in such situations, which has been shown to be linked, in part, to the voice differences between target and masking speakers (e.g., Brungart, 2001; Festen and Plomp, 1990; Stickney *et al.*, 2004). In those studies, target recognition

scores were found to improve when the gender of the masking speaker was different from that of the target, compared to the baseline conditions where the target and masker were either the same speaker or were of the same gender.

Such voice differences between speakers can be decomposed largely along two dimensions, namely, the voice pitch and the vocal tract length (VTL). The voice pitch is the perceptual correlate of the fundamental frequency ( $F_0$ ) that arises from the glottal pulse rate, while the VTL dimension is correlated with body size, and hence gives cues to the size of the speaker (Evans *et al.*, 2006; Fitch and Giedd, 1999; Ives *et al.*, 2005; Smith and Patterson, 2005). Manipulating both of these cues together was found to elicit a change in perceived speaker gender (Hillenbrand and Clark, 2009; Skuk and Schweinberger, 2014; Smith and Patterson, 2005). In addition, increasing the difference in  $F_0$  (Assmann and Summerfield, 1990; Başkent and Gaudrain, 2016; Brox and Nooteboom, 1982; Darwin *et al.*, 2003; Drullman and Bronkhorst, 2004; Lee and Humes, 2012), VTL (Başkent and Gaudrain, 2016; Darwin *et al.*, 2003), or both (Başkent and Gaudrain, 2016; Darwin *et al.*, 2003; Vestergaard *et al.*, 2009) between target and masking speakers was shown to yield a systematic increase in target sentence identification scores for NH listeners. On the other hand, no release from masking for CI users was observed when either  $F_0$  (Pyschny *et al.*, 2011; Stickney *et al.*, 2007), VTL (Pyschny *et al.*, 2011), or both (Pyschny *et al.*, 2011) were varied between target and masking speakers, or when completely

<sup>a)</sup>Portions of this work were presented in “Effect of frequency allocation on vocal tract length perception in cochlear implant users,” Conference on Implantable Auditory Prostheses (CIAP’ 15).

<sup>b)</sup>Also at: University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Groningen, The Netherlands. Electronic mail: n.el.boghdady@umcg.nl

<sup>c)</sup>Also at: University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Groningen, The Netherlands.

<sup>d)</sup>Also at: University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Groningen, The Netherlands and CNRS UMR 5292, INSERM U1028, Lyon Neuroscience Research Center, University Lyon, Lyon, France.

different speakers were used as target and masker (Stickney *et al.*, 2004).

The inability of CI users to benefit from  $F0$  and VTL differences may arise from their abnormal perception of these two cues. For example, not only do CI users demonstrate poor sensitivity to differences in both  $F0$  and VTL compared to NH listeners (Gaudrain and Başkent, 2018), but they are also unable to use the latter to correctly judge a speaker's gender (Fuller *et al.*, 2014; Meister *et al.*, 2016).

This reduced sensitivity to  $F0$  and VTL differences may be attributed to the poor spectral resolution in the implant (Friesen *et al.*, 2001; Fu *et al.*, 1998; Henry and Turner, 2003; Winn *et al.*, 2016), which is likely more detrimental to VTL cues than to  $F0$  (Gaudrain and Başkent, 2015). This is because VTL information is mainly represented by the formant peaks in the spectral envelope of the signal (Chiba and Kajiyama, 1941; Fant, 1960; Lieberman and Blumstein, 1988; Müller, 1848; Stevens and House, 1955), as opposed to  $F0$  cues, which were shown to be encoded both in the temporal envelope and the corresponding place of stimulation along the cochlea (e.g., Carlyon and Shackleton, 1994; Licklider, 1954; Oxenham, 2008).

Effective spectral resolution in the implant can be dictated by a number of factors, including the amount of channel interaction, the effective number of spectral channels, and the resolution of the frequency band partitioning map (for a review, see Başkent *et al.*, 2016). Channel interaction occurs due to current spread between neighbouring electrodes (e.g., Boëx *et al.*, 2003; De Balthasar *et al.*, 2003; Hanekom and Shannon, 1998; Shannon, 1983; Townshend and White, 1987), which results in reducing the number of effective spectral channels. It was suggested that CI users have no more than 8 effective spectral channels, as opposed to NH listeners, who have up to 20–24 effective spectral channels under vocoded conditions (Friesen *et al.*, 2001; Qin and Oxenham, 2003). Both increased channel interaction and reduced number of effective channels were found to negatively impact not only speech and phoneme perception (e.g., Friesen *et al.*, 2001; Fu and Shannon, 2002; Qin and Oxenham, 2003), but also VTL sensitivity under vocoder simulations (Gaudrain and Başkent, 2015).

The frequency band partitioning map is used to quantize the spectral information received by the implant into a number of contiguous channels. The information in each channel is usually delivered to a separate electrode in the stimulating array, which determines the resolution (number of electrode channels) dedicated to the specified frequency range. To minimize trauma while maintaining sufficient stimulation of surviving auditory nerve fibres, electrode arrays are seldom inserted more than 2.6 rounds into the cochlea (Skinner *et al.*, 2007). This means that the frequency corresponding to the location of the most apical electrode falls between about 250 Hz and 870 Hz, depending on the cochlear dimensions, electrode array length, and insertion depth (Franke-Trieger and Mürbe, 2015; Skinner *et al.*, 2007). Consequently, if the frequency partitioning map fully matches the frequencies corresponding to electrode locations, low-frequency information important for speech intelligibility would be lost (Başkent and Shannon, 2004), especially for cases in which

the most apical electrode location corresponds to around 800 Hz. Conversely, if the full typical range of the frequency partitioning map (from around 200 Hz to 8 kHz) is allocated to the electrodes, speech intelligibility would also be impaired (Başkent and Shannon, 2004). This inevitably yields a frequency mismatch between the frequencies received by the implant and those corresponding to actual places of stimulation along the cochlea.

The degree of mismatch differs across CI users due to the variability in cochlear dimensions (Avci *et al.*, 2014; van der Marel *et al.*, 2014) and electrode array designs and their corresponding insertion depths (Finley *et al.*, 2008). However, in clinical practice, the frequency band partitioning maps are seldom customized for each individual CI user (Fitzgerald *et al.*, 2013; Landsberger *et al.*, 2015; Tan *et al.*, 2017; Venail *et al.*, 2015). A number of studies have suggested optimizing the frequency band partitioning map in implant processing to help alleviate the negative effects of frequency mismatch, and hence improve performance on a number of tasks, such as melodic pitch perception (Di Nardo *et al.*, 2011; Omran *et al.*, 2011), phoneme recognition (Fu and Shannon, 1999a, 2002; Leigh *et al.*, 2004; McKay and Henshall, 2002), and speech intelligibility (Fitzgerald *et al.*, 2013; Grasmeyer *et al.*, 2014; McKay and Henshall, 2002).

The aim of the present study was to assess the impact of frequency mismatch and band partitioning on VTL sensitivity, using acoustic vocoder simulations of CI processing with NH listeners. These vocoder simulations (Dudley, 1939; Fu and Shannon, 1999b; Gaudrain and Başkent, 2015; Shannon *et al.*, 1995; Shannon *et al.*, 1998) were used to better specify the parameters in each frequency mismatch and band partitioning setup, as these would be difficult to control for in actual CI users (Fitzgerald *et al.*, 2013). Just-noticeable-differences (JNDs) for VTL were collected as a measure of sensitivity following the protocol described by Gaudrain and Başkent (2015, 2018).

Frequency mismatch and band partitioning were studied by addressing three research questions, to each of which a separate experiment was dedicated. The first research question, addressed in experiment 1, was whether simulating a simple frequency mismatch by introducing a shift between the vocoder analysis and synthesis filters would affect the VTL JNDs. This was motivated by the findings of Shannon *et al.* (1998), which showed that simulated frequency shift impaired vowel recognition; a stimulus type that likely has cues that are affected in a similar manner to those of VTL. This is because the representation of both vowel differences and VTL cues lies in the structure of formant frequencies. Thus, the hypothesis for this experiment was that the larger the simulated mismatch (shift) between the analysis and synthesis filters, the worse the VTL sensitivity would become.

The second research question, addressed in experiment 2, was whether the choice of frequency band partitioning would affect VTL JNDs when no frequency mismatch is present. This was crucial to test, because if band partitioning had an effect on VTL JNDs, then this would imply that optimal band partitioning may have the potential to mitigate the detrimental effects of frequency mismatch on VTL sensitivity. The hypothesis was that a band partitioning scheme,

which dedicates a larger number of bands to the lower frequency components (higher spectral resolution), would better transmit formant frequencies, where VTL cues are encoded. Hence, this band partitioning scheme is expected to improve VTL sensitivity compared to a band partitioning with a lower spectral resolution at the lower frequencies. A similar finding was reported by Shannon *et al.* (1998) such that higher spectral resolution near the lower frequencies yielded better vowel recognition scores.

The final research question, addressed in experiment 3, was related to the combined effect of both frequency mismatch and band partitioning in a more realistic simulation of CI processing. This was done to investigate whether indeed a frequency partitioning map with sufficient spectral resolution in the lower frequencies would help preserve VTL cues, irrespective of the severity of the frequency mismatch.

## II. GENERAL METHODS

### A. Stimuli

The stimulus design was identical to that previously used by Gaudrain and Başkent (2015). Speech material was taken from the Nederlandse Vereniging voor Audiologie (NVA) corpus (Bosman and Smoorenburg, 1995), which is a collection of lists of meaningful monosyllabic consonant-vowel-consonant (CVC) Dutch words uttered by a female speaker. Sixty-one consonant-vowel (CV) syllables, with a duration between 142 ms and 200 ms, were manually extracted from the list of NVA words. Co-articulation between the vowel and final consonant in the original CVC file was minimized by applying a cosine offset ramp of 60 ms to the end of the extracted syllable. Moreover, a cosine onset ramp of 5 ms was applied to the beginning of the syllable to make it sound more natural and to avoid spectral splatter. The finalised CV syllable list consisted of combinations of the consonants [b, d, f, k, l, m, n, p, r, s, t, v, x, z] and vowels [ɛ, a, e, o, ɤ, α, i, u, ɔ, ɪ], and was equalised in root-mean-square (rms) intensity. The duration of each syllable was normalised to 200 ms using STRAIGHT (Kawahara and Irino, 2005).

For all three experiments, the stimuli in each trial were created by randomly selecting 3 different CV syllables from the available list of 61 syllables and stringing them together with a 50 ms inter-syllable interval to form a triplet. In each trial, a new triplet of syllables was formed, but within a trial, the same triplet of syllables was presented three times with a silent gap of 250 ms between each presentation. Only one of these three presentations had a different VTL (processed using STRAIGHT) relative to the other two identical presentations, while the average  $F_0$  over each presentation was held constant. Hence, the procedure was an adaptive “odd-one-out,” i.e., a three-interval, three-alternative forced choice task (3I-3AFC), where the participant had to select the interval (triplet) that had a different VTL relative to the other two. All three triplets were resynthesized by STRAIGHT, even when  $F_0$  and VTL were not changed relative to the original female voice.

Figure 1 shows how VTL was manipulated in this study, where  $\Delta VTL$  is the ratio expressed in semitones (st) between VTL of the synthesized speaker and that of the original

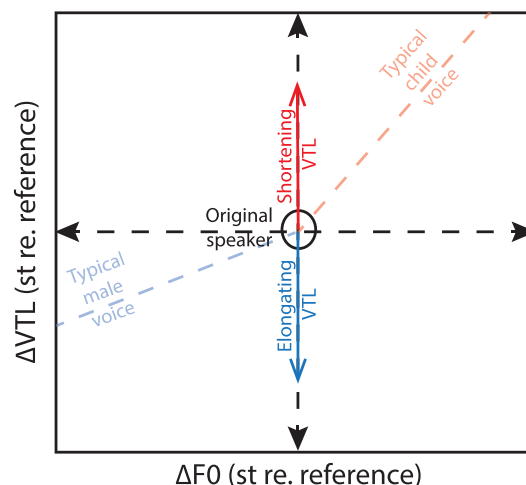


FIG. 1. (Color online) VTL manipulations shown along the  $F_0$ -VTL plane in reference to the original female voice at the origin of the plane. For further clarity, typical male and children voices are also marked on the same plane.

speaker. Shortening (elongating) VTL translates into stretching (compressing) the spectral envelope of the signal relative to the original. Thus, in order to realize changes in VTL, STRAIGHT manipulates the spectral envelope of the synthesized signal in relative changes with respect to the original (Patterson and Smith, 2003; Smith *et al.*, 2005).

### B. Apparatus

All three experiments were conducted in a sound-attenuated booth, and stimuli were presented through HD600 headphones (Sennheiser GmbH and Co., Wedemark, Germany) via an AudioFire4 soundcard (Echo Digital Audio Corp, Santa Barbara, CA) connected to a DA10 D/A converter (Lavry Engineering, Poulsbo, WA) through Sony/Philips Digital Interface. The output from this setup was calibrated to a level of 65 dB sound pressure level (SPL) (except for experiment 1, which was calibrated to 60 dB SPL) using a KEMAR head and torso assembly Type 45BA (G.R.A.S. Sound and Vibration, Holte, Denmark). All signal processing and stimulus presentations were performed in MATLAB R2014b (The Mathworks, Natick, MA) using a sampling frequency of 44.1 kHz, and all data analyses were done in R (R Core Team, 2014).

### C. Vocoder simulations

Noise-band vocoders (Dudley, 1939; Shannon *et al.*, 1995) were used in this study to acoustically simulate CI processing. The frequency-to-electrode allocation map in a typical CI processing pathway was modeled by the vocoder analysis filters. The frequency mismatch in the implant was modeled by the differences in frequency band setups between the vocoder analysis and synthesis filters (e.g., as was done by Shannon *et al.*, 1998). Vocoding was implemented by extracting the temporal envelope from each analysis filter band by half-wave rectification and low-pass filtering at a cutoff of 300 Hz using a zero-phase, fourth-order Butterworth filter. These envelopes were used to modulate a white noise carrier signal, which were then filtered by

the set of synthesis filters after modulation. The vocoded signal was obtained by summing the modulated output from all frequency bands. Figure 2 depicts the analysis and synthesis filter settings for each experiment.

### 1. Analysis filters

The analysis bandpass filters were implemented using zero-phase Butterworth filters, whose order (slope) differed across experiments. In experiment 1, 12 filter bands of fourth- and eighth-order were used to simulate the effect of channel interaction. Both analysis and synthesis filters were given the same filter order for a given condition. This choice of filter orders was based on data from [Gaudrain and Başkent \(2015\)](#), which showed that shallower filters, simulating larger channel interaction, yielded VTL JNDs that were close to those obtained from actual CI users ([Gaudrain and Başkent, 2018](#)). It is expected that frequency shift might play a larger role with sharper filters than with shallower filters because shallow filters effectively become more similar to each other, which should manifest as an interaction effect between filter order and frequency shift. In experiments 2 and 3, 16 analysis filter bands of 12th-order were used instead because pilot data revealed that 4th- and 8th-order filters, when combined with the synthesis filter models used in experiment 3, yielded unrealistically large VTL JNDs compared to those of actual CI users ([Gaudrain and Başkent, 2018](#)).

The parameters for band partitioning were determined based on previous work on optimizing frequency band partitioning for a range of tasks (e.g., [Başkent and Shannon, 2004, 2005](#); [Fitzgerald et al., 2013](#); [Fu and Shannon, 1999b, 2002](#); [McKay and Henshall, 2002](#); [Shannon et al., 1998](#)). The maps used in those studies (replotted in the [Appendix](#)) varied between either a logarithmic-like (Greenwood-like) partitioning or a purely linear partitioning. The Greenwood formula, reproduced as Eq. (1) ([Greenwood, 1990](#)), describes the logarithmic-like relationship between a given location,  $x$  (in millimetres), along the human basilar membrane relative to the average length of the cochlea,  $C$ , and its corresponding tonotopic frequency,  $F$ , in Hertz,

$$F_i = A(10^{(C-x)/a} - k). \quad (1)$$

The parameters in Eq. (1) were set to  $A = 165.4$ ,  $a = 0.06$ , and  $k = 0.88$  based on those provided by [Greenwood \(1990\)](#) for a human cochlea. The average cochlear length,  $C$ , was set to the typical value of 35 mm (e.g., as was done by [Başkent and Shannon, 2004, 2005](#); [Fu and Shannon, 1999b](#)). The subscript  $i$  refers to the  $i$ th cut-off frequency.

VTL modification affects all frequencies by the same ratio, i.e., it is a pure translation on a log-frequency axis. Because the natural frequency-place relationship is not perfectly logarithmic (as shown by the “ $-k$ ” in Greenwood’s formula), a VTL shift does not result in a uniform translation in terms of place of stimulation. Hence, frequency mismatch in the implant can be expected to impair VTL cues, which may be addressed by adjusting the frequency partitioning map. Compared to a logarithmic-like or Greenwood partitioning, linearly partitioned maps have fewer channels dedicated to

the lower frequencies, hence, would be expected to smear the formant peaks in that frequency range, leading to a distortion in VTL cues. Thus, in this study, a partitioning based on the Greenwood formula and a linear partitioning were chosen for the analysis filters based on the literature. Additionally, two more maps were chosen based on what is available in actual clinical devices in order to have a measure of how well these maps can convey VTL cues in simulation. One of these clinical maps was based on the Advanced Bionics (AB) HiRes 90K map (Stäfa, Switzerland/Valencia, CA), and the other on Frequency Table 22 from Cochlear (Macquarie University Sydney, NSW, Australia).

The overall frequency range of the analysis filters of the frequency partitioning maps differed across experiments. In experiment 1, the analysis filters covered the range between 150 Hz and 7000 Hz and were partitioned into 13 bands in equal simulated cochlear distance according to the Greenwood function ([Gaudrain and Başkent, 2015](#)). In experiments 2 and 3, the analysis filters covered the frequency range from 250 Hz to 8700 Hz. This change was made so that all maps eventually used in experiment 3 would cover a frequency range similar to the standard map assigned to the electrode array model used for designing the synthesis filters (see Sec. IIC2). In experiment 2 the analysis filters were partitioned once according to Greenwood (as was done in experiment 1) and once using linear spacing. The linear map was obtained by taking 17 linearly spaced points along the frequency scale between 250 Hz and 8700 Hz. In experiment 3, the same Greenwood and linear maps defined in experiment 2 were used, and the HiRes and Cochlear maps were added. The HiRes 90K implant model was chosen because it is rather common, and thus would serve as a reasonable simulation. This map has 17 cut-off frequencies (16 channels) between 250 Hz and 8700 Hz. Because the Cochlear map has 22 channels with 23 cutoffs between 188 Hz and 7938 Hz, it was compressed to 16 channels by linearly interpolating the cut-off frequencies between 188 Hz and 7938 Hz at 17 equally spaced points. This was done to prevent potential advantages in JNDs that may result from a larger number of channels (and thus a higher spectral resolution).

### 2. Synthesis filters

Across experiments, frequency mismatch was simulated by introducing differences between the analysis and synthesis filters. In experiment 1, the synthesis filters were derived from the analysis filters by basally shifting all the frequencies by 0, 2, 4, and 6 mm relative to a 35-mm-long cochlea ([Başkent and Shannon, 2005](#); [Finley et al., 2008](#); [Fitzgerald et al., 2013](#); [Fu and Shannon, 1999b](#)), as shown in panel 1 of Fig. 2. In experiment 2, because only the effect of frequency partitioning without mismatch was of interest, the synthesis filters were kept identical to the analysis filters under each condition (see panel 2 of Fig. 2). In experiment 3, the synthesis filters were designed to more closely model the maps in realistic CI systems, using dimensions from actual implants. These synthesis bandpass filters were created using

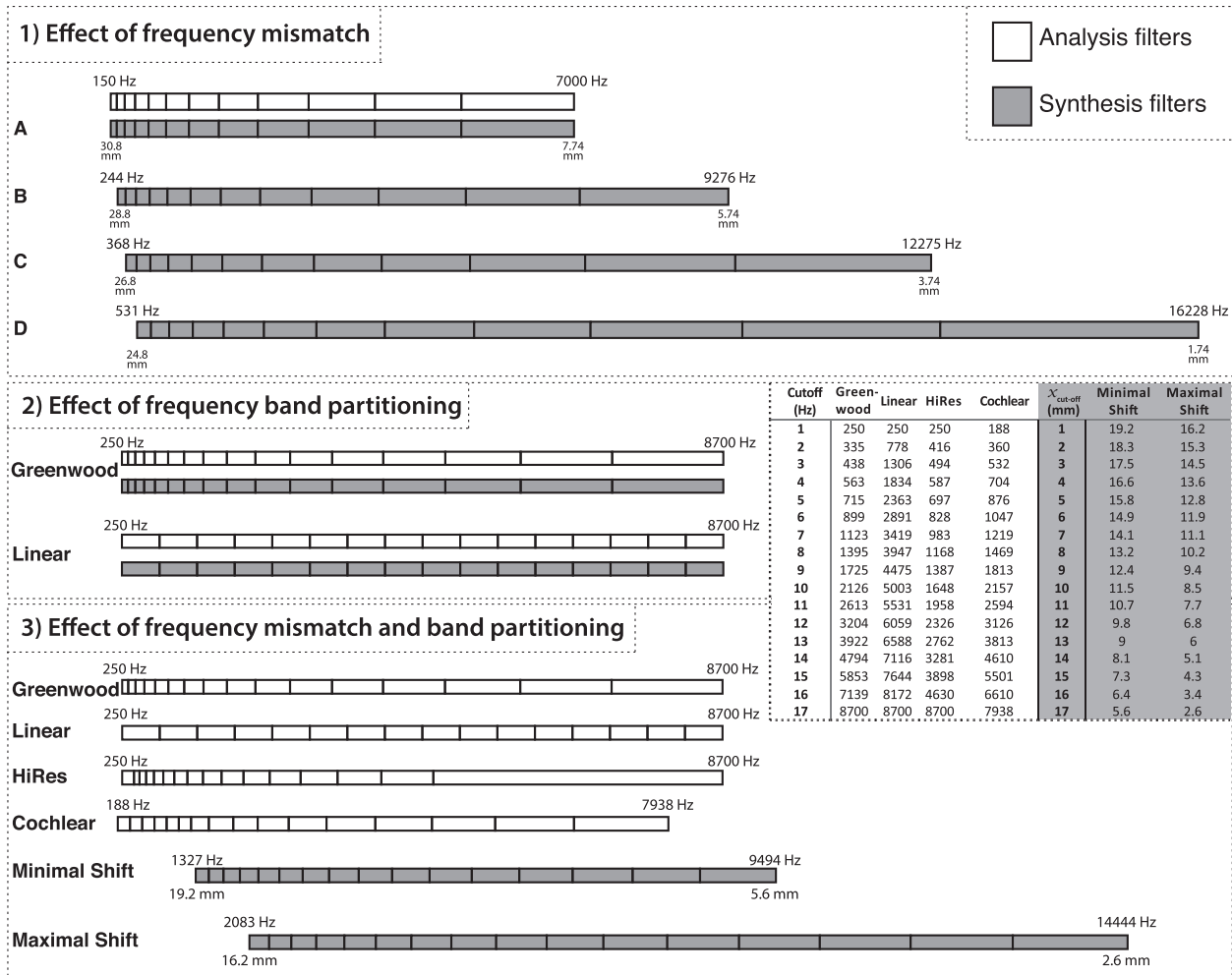


FIG. 2. Vocoder analysis (white bands) and synthesis (grey bands) filters shown for all three experiments, as partitioned along frequency. Cut-off frequencies are shown only for the most apical and most basal bands, along with their corresponding locations in millimetres, where applicable, relative to the base of a 35-mm-long cochlea. (1) Vocoder setup for experiment 1, where the frequency mismatch was produced by systematically shifting the synthesis filters basally from the analysis filters by (A) 0 mm, (B) 2 mm, (C) 4 mm, (D) 6 mm. (2) Vocoder setup for experiment 2, where band partitioning was introduced in the analysis filters, while the cut-off frequencies of the synthesis filters were identical to those of the analysis filters under a given condition. (3) Vocoder setup for experiment 3, where frequency mismatch and band partitioning were combined.

16 zero-phase fourth-order Butterworth filters to account for the effect of spread of excitation, with centre frequencies computed via Eq. (1),

$$x_i = x_0 + d(i - 1), \quad i = 1, 2, \dots, 16. \quad (2)$$

For the synthesis filters,  $x_i$  in Eq. (1) was computed as shown in Eq. (2) (Fu and Shannon, 1999b), and represents the position corresponding to the centre of the  $i$ th simulated electrode along the 35-mm-long basilar membrane.  $x_0$  represents the position of the first electrode in the simulated array from the base of the cochlea,  $d$  represents the inter-electrode spacing centre-to-centre, and  $i$  represents the simulated electrode number.

The parameters for this equation were based on the dimensions of the 24.5-mm-long AB HiFocus Helix electrode array (Sylmar, 2005), which belongs to a family of electrode models under the HiRes 90K implant. The AB HiFocus Helix array was specifically chosen here because its dimensions yield a model that is comparable to the one used by Fu and Shannon (1999b), and thus gives a reference to

which the current model proposed here can be compared. Two possible electrode array insertion depths were determined from the locations of the proximal and distal markers; inserting the electrode array up to the proximal marker yields an insertion depth of about 21.5 mm from the base of the cochlea, while inserting it up to the distal marker yields an insertion depth of around 18.5 mm (Sylmar, 2005). The position of the first simulated electrode,  $x_0$ , was computed by subtracting the length of the active contact area of the array (15.5 mm), where the stimulating electrodes lie, from these two possible insertion depths. This yielded values for  $x_0$  of either 6 mm for an array inserted up to the proximal marker, or 3 mm for an array inserted up to the distal marker. These two conditions are referred to as *minimal shift* and *maximal shift*, respectively, in the rest of this paper. In Eq. (2), the inter-electrode spacing,  $d$ , was set to 0.85 mm, as defined in the surgical manual (Sylmar, 2005).

The cut-off frequencies of the synthesis filters ( $x_{\text{cutoff}}$  in Fig. 2), were defined by the frequencies corresponding to the mid-distance point between the electrode centres (computed

using the inter-electrode spacing,  $d$ ). The values of  $x_{\text{cutoff}}$  are shown in millimetres in the table provided in Fig. 2.

#### D. Procedure for measuring VTL JNDs

Each JND for a given run was obtained using a two-down one-up adaptive procedure, yielding 70.7%-correct on the psychometric function (Levitt, 1971). The initial trial started at a VTL difference of 12 st between reference and target triplets along either VTL manipulation type (i.e., elongating or shortening VTL). The reference voice was always that of the original female speaker. After each two successive correct responses, the absolute VTL difference between the reference and target triplets decreased by a step size of 4 st. After a single incorrect response, the VTL difference was increased by the same step size. If the VTL difference became smaller than twice the step size, the step size was reduced by a factor of  $\sqrt{2}$ . The run terminated after eight reversals, and the JND was calculated as the mean VTL difference, in st, between the target and reference triplets obtained in the last six reversals. The run stopped automatically after 150 trials if the algorithm had not converged by then, and the measurement was discarded.

Training was provided for 15 min at the beginning of the first session with the purpose of familiarizing participants with the test procedure. In the training phase, the two VTL manipulations were used, in addition to two vocoder settings, forming a total of four conditions. These four conditions were presented in a pseudo-random order, with visual feedback showing the participant whether the interval they selected was correct or not. This type of feedback was also provided during actual testing. Each training run was programmed to end after only six trials, irrespective of whether the adaptive procedure converged or not.

### III. EXPERIMENT 1: EFFECT OF FREQUENCY SHIFT AND FILTER ORDER ON VTL JNDs

The effect of frequency mismatch on VTL JNDs in vocoder simulations was investigated by introducing a place shift between the analysis and synthesis filters of the vocoder. Because channel interaction [simulated as vocoder filter order (slope)] was shown in previous simulation studies to influence both vowel identification (Shannon *et al.*, 1998) and VTL JNDs (Gaudrain and Başkent, 2015), it was also investigated in this experiment for possible interactions with frequency shift. The expectations were that VTL JNDs would worsen as the frequency shift and simulated channel interaction increased.

#### A. Methods

##### 1. Participants

Fifteen NH listeners, aged 19–40 years old ( $\mu = 25.1$  yr,  $\sigma = 5.9$  yr), participated in this experiment. Amongst the 15 participants, 12 had already taken part in similar experiments (Gaudrain and Başkent, 2015). Their audiometric thresholds were tested at octave frequencies between 250 Hz and 8000 Hz and found to be all below 20 dB hearing level (HL). All participants had no history of hearing disorders, dyslexia,

or attention deficit hyperactivity disorder, were generally in good health, and were either native Dutch speakers, or had Dutch as one of the languages used in their daily childhood environment. Participants provided signed informed consent prior to data collection, and the entire study protocol was approved by the ethics committee of the University Medical Center Groningen (METc 2012.392). Finally, all participants received an hourly wage for their participation, in accordance with the department guidelines.

#### 2. Procedure

The procedure was as described in Sec. II (General Methods), with the following additional details. A total of 16 experimental conditions were administered: 2 types of VTL manipulations (elongating and shortening VTL)  $\times$  2 filter orders (4, 8)  $\times$  4 frequency shift values (0, 2, 4, 6 mm). Each condition was repeated twice for a total of 32 runs, which were randomly split into two sessions of 16 runs each. Each session lasted for 2 h and was conducted on a separate day.

### B. Results and discussion

Figure 3 shows the distribution of VTL JNDs across all participants as a function of frequency shift and filter order. The horizontal dashed line in Fig. 3 shows the typical VTL difference between a male and a female voice as used for the gender categorization experiment by Fuller *et al.* (2014). For the sharper filters (eighth-order), when the analysis and synthesis filters were aligned, most of the participants in the current study were able to discriminate VTL values that

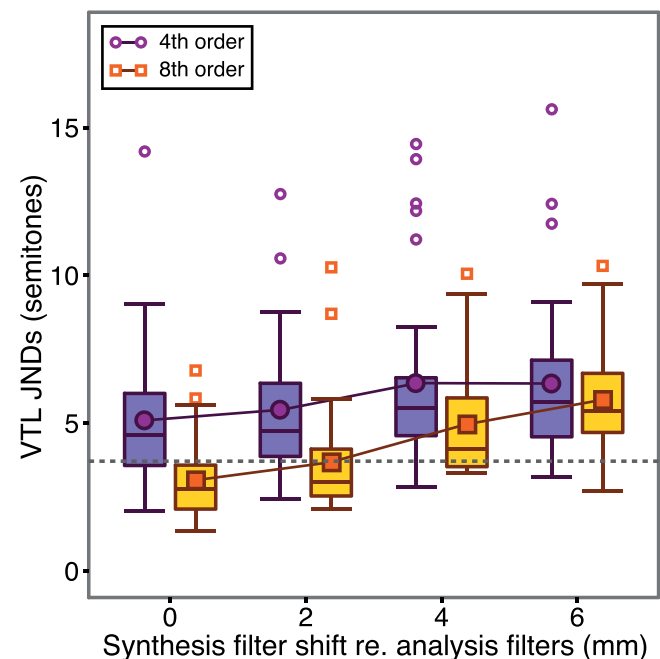


FIG. 3. (Color online) VTL JNDs shown as a function of filter order and frequency shift. The boxes extend from the lower to the upper quartile, and the middle line shows the median. The filled symbols (circle and square) show the means for fourth- and eighth-order filters, respectively. The whiskers show the range of the data within 1.5 times the inter-quartile range (IQR). The empty symbols show the individual data outside of 1.5 times IQR. The horizontal dashed line represents the difference in VTL that was used to represent a typical difference between the male and female voices in Fuller *et al.* (2014).



corresponded to this typical male-female VTL difference. This means that the VTL cue should be available to them to perform a gender categorization task. However, when the synthesis filters were shifted by 6 mm in the basal direction, almost all the participants' JNDs became larger than this typical male-female VTL difference. With such a shift, they would thus become unable to use the VTL cue for gender categorization purposes.

A three-way repeated-measures analysis of variance (ANOVA) was performed on the log-transformed JNDs, with VTL manipulation (elongating and shortening), filter order, and frequency shift as repeated factors. The JNDs were log-transformed to improve the homoscedasticity of the data set and because the adaptive procedure is such that only positive threshold values can be reached, and the step size evolves logarithmically. The VTL manipulation was found to have a small but significant effect on the JNDs [ $F(1,14) = 5.71$ ,  $p = 0.03$ ,  $\eta_G^2 = 0.02$ ]: the average JND measured starting from longer VTLs was 5.21 st, while it was 4.67 st when starting from shorter VTLs. The effect of frequency shift was found to be significant [ $F(3,42) = 30.56$ ,  $p < 0.0001$ ,  $\eta_G^2 = 0.13$ ]: the larger the shift between analysis and synthesis filters, the worse the JNDs were. The order of the filters also significantly affected the JNDs [ $F(1,14) = 26.54$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.11$ ]: sharper filters yielded smaller JNDs, consistent with the findings of Gaudrain and Başkent (2015). This effect interacted with the frequency shift [ $F(3,42) = 7.85$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.03$ ]: for a shift of 6 mm, the difference between the mean JNDs for the two filter orders was 0.4 st, while when no shift was introduced, the difference between the two filter orders was 2.0 st. This indicates that the broader the channels, the less effect the frequency shift has on VTL JNDs (but note the small effect size). All other interactions were non-significant ( $p > 0.10$ ).

Systematically increasing the frequency shift led to a decrease in the sensitivity to VTL differences. This finding is compatible with the hypothesis that introducing a frequency shift can hinder access to VTL cues, and is in line with the findings reported by Başkent and Shannon (2004), Fu and Shannon (1999b), and Shannon *et al.* (1998), where frequency shifts largely reduced vowel recognition scores in those studies. These results thus suggest that the frequency shift that occurs in implants may contribute to the poor VTL JNDs observed in implant users.

Figure 4 shows how a VTL difference is represented along the cochlear partition depending on the degree of shift introduced between the vocoder analysis and synthesis filters. When the difference is represented as a function of log-frequency (lower left panel), it appears that the cues are compressed in frequency, which is a tempting explanation as to why the sensitivity was lower in the 6-mm shift case. However, when expressed as a function of equivalent rectangular bandwidth (ERB) number (lower right panel), the difference between the two vocoder conditions becomes minimal. In other words, while physical representations of the signals resulting from the two extreme shift conditions appear to be quite different, basic estimates of the perceptual representations do not display such large differences. It thus seems unlikely that the poor sensitivity to VTL differences

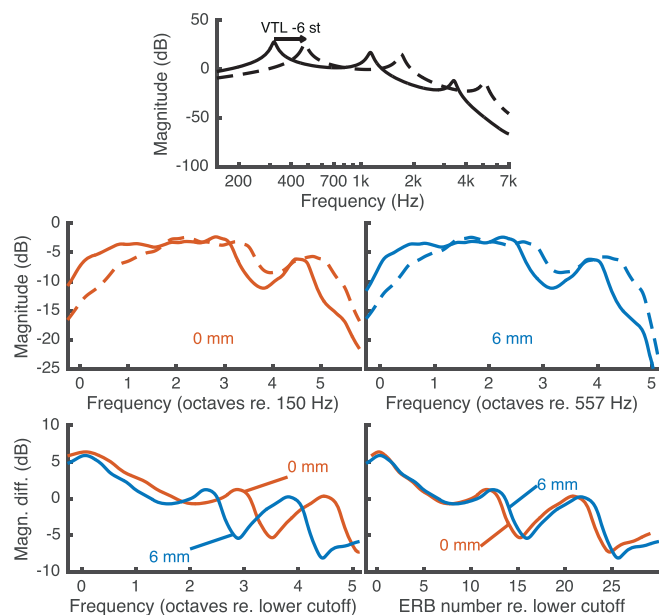


FIG. 4. (Color online) Representation of a VTL difference through matched and shifted analysis and synthesis filters. (Top) Schematic spectra of an artificial, three-formant vowel. The solid line represents the original vowel, and the dashed line represents the same vowel produced with a VTL 1.5 times shorter (equivalent to a  $-6$  st shift). (Middle) Magnitude spectra of the vocoded versions of the same vowels for the eighth-order vocoder, with a frequency shift of 0 mm (left) and 6 mm (right). Note that the frequency axis is expressed in octaves relative to the lower cutoff of the first synthesis filter. (Bottom) These panels show the difference between the solid and dashed lines in the middle row, thus, illustrating how the VTL difference is represented for the two vocoder conditions. The left panel shows the difference as a function of octave frequency relative to the lower cut-off frequency of the first synthesis filter (which is different for 0 mm and 6 mm shift vocoders). The right panel shows the same but with the frequency expressed in equivalent rectangular bandwidth (ERB) number.

observed with 6-mm shift could be explained by a spectral distortion of the VTL cues induced by the shift.

A perhaps more plausible explanation for these results is that the 6-mm shift condition presents speech in an unusual frequency region, where NH listeners may have never been exposed to VTL differences before, unlike the case for the frequency region involved in the 0-mm shift condition. This would be consistent with the findings of Ives *et al.* (2005) who reported VTL JNDs that were largest for voices with formants falling in the higher frequencies. If this is indeed the case that lack of prior exposure to frequency-shifted speech can explain the present lack of sensitivity to VTL differences in the 6-mm shift condition, then one might venture that training could improve VTL discrimination performance. However, Massida *et al.* (2013) measured sensitivity to voice gender difference in CI users over 18 months after implantation and observed no improvement over this period of time. Thus, if frequency shift contributes to the reduced VTL JNDs observed in CI users, it seems that this hindrance may not be easily alleviated by unsupervised exposure to speech sounds.

One potential limitation to the above conclusion is that, in the condition with the largest shift, the upper channels correspond to a frequency region that was not assessed in the audiometric test undertaken with the participants. While NH was only assessed up to 8 kHz, the two most basal synthesis

filters for a shift of 6 mm spanned from 9.6 to 12.5 kHz, and from 12.5 to 16.3 kHz. It is thus possible that these channels were not clearly audible to the participants. However, because this lack of audibility only concerns two channels that are least likely to carry crucial VTL information, it seems relatively unlikely that audibility alone could explain the effect of frequency shift observed here. Nonetheless, this concern was addressed in experiment 3, such that audiometric thresholds above 8 kHz were measured for all participants.

Moreover, such a limitation would not apply to actual CI users, however, other aspects of the vocoders used in this first experiment might hinder the generalisation of these findings to electric hearing. First, the analysis filterbank used in this experiment has channels that are equidistant in terms of stimulation place along the basilar membrane. In contrast, the filterbanks used in commercial CI processors do not follow this partitioning. In addition, while permitting the systematic assessment of the effect of frequency shift on VTL sensitivity, the vocoders used in this experiment do not accurately mimic how commercial CIs deliver spectral information. This was also addressed in experiment 3, where a more realistic vocoder setup was used.

In this experiment, while the effect of frequency shift on VTL sensitivity was investigated, the effect of band partitioning was not assessed. Hence, the effect of band partitioning on VTL JNDs was studied in experiment 2.

## IV. EXPERIMENT 2: EFFECT OF FREQUENCY BAND PARTITIONING ON VTL JNDS

### A. Rationale

The aim of this experiment was to investigate the effect of frequency band partitioning on VTL JNDs in vocoder simulations of CI processing. VTL changes are realized as a shift in all formant peaks of the spectral envelope of the signal by the same amount on a log-frequency axis. This means that in order to properly convey such subtle shifts in spectral peaks, the frequency band partitioning in the implant needs to have a sufficiently high resolution in the frequency region where formant peaks are usually represented. Thus, the proposed hypothesis in this experiment is that a filterbank with more channels dedicated to frequencies lower than 3 kHz, where the first formants are encoded, is expected to yield smaller VTL JNDs, compared to a map with fewer channels in that frequency region. For this reason, two such partitioning maps were tested in this experiment, and assigned as the analysis filters: the Greenwood map, which has a higher resolution for frequencies below about 3 kHz, and the linear map, which has a lower resolution in this frequency region (see panel 2 of Fig. 2). Here, only the effect of frequency partitioning was studied; the synthesis filters were an exact copy of the analysis filters in each condition to remove any effects of frequency mismatch.

### B. Methods

#### 1. Participants

Using same inclusion criteria as in experiment 1, 16 NH young adults (age: 18–30 yr,  $\mu = 22.6$  yr,  $\sigma = 3.2$  yr), different

than those recruited for experiment 1, participated in this experiment. One participant did not return to complete the experiment; their data were excluded from the analyses, resulting in a total of 15 participants (age: 18–30 yr,  $\mu = 22.7$  yr,  $\sigma = 3.3$  yr), whose data were analysed.

## 2. Procedure

The procedure was as described in Sec. II (General Methods), with four administered experimental conditions. These were composed of the 2 types of VTL manipulations (elongating and shortening VTL)  $\times$  2 frequency band partitioning maps (Greenwood and linear).

## C. Results and discussion

Figure 5 shows the JNDs obtained from the Greenwood and linear partitioning maps tested in this experiment for elongating or shortening VTL.

A two-way repeated measures ANOVA was applied on the log-transformed JNDs, with frequency partitioning map and VTL manipulation as repeated factors. Confirming the hypothesis, the analysis revealed that the linear map was indeed significantly worse than the Greenwood map by about 3.35 st on average [ $F(1,14) = 85.97$ ,  $p < 0.0001$ ,  $\eta_G^2 = 0.31$ ]. A pairwise  $t$ -test with false discover rate (FDR) correction for multiple comparisons (Benjamini and Yekutieli, 2001) was applied to compare both maps for each VTL manipulation individually. This also revealed that the Greenwood map was significantly better than the linear map for both elongating

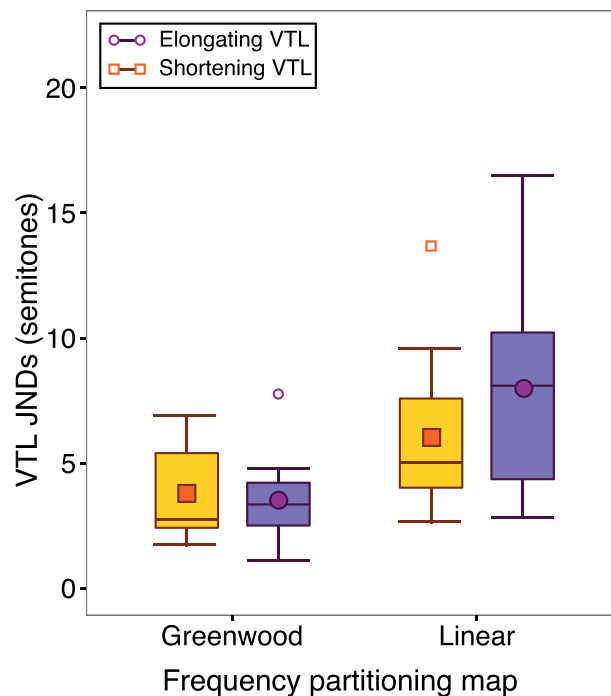


FIG. 5. (Color online) VTL JNDs shown as a function of frequency partitioning map and VTL manipulation. The boxes extend from the lower to the upper quartile, and the middle line shows the median. The filled circles and squares show the means for elongating and shortening VTL, respectively. Hollow symbols represent outliers. The details for the boxplot are as described in Fig. 3.

$[t(14) = 6.32, p_{\text{FDR}} < 0.0001, \delta = 4.47 \text{ st}]$  and shortening VTL  $[t(14) = 8.35, p_{\text{FDR}} < 0.0001, \delta = 2.24 \text{ st}]$ .

The intriguing finding was that the frequency partitioning maps affected the JNDs differently depending on the VTL manipulation type, as indicated by the significant interaction effect between these two factors  $[F(1,14) = 5.4, p = 0.036, \eta_G^2 = 0.029]$ . With the Greenwood map, participants were equally sensitive to longer and shorter VTLs  $[t(14) = 0.49, p_{\text{FDR}} = 0.63, \delta = 0.27 \text{ st}]$ , but with the linear map, participants were more sensitive to shorter VTLs than longer VTLs  $[t(14) = 2.29, p_{\text{FDR}} = 0.050, \delta = 1.96 \text{ st}]$  (but note the small effect size and the borderline significant effect). This behaviour is expected for the linear map because it has a smaller number of channels for frequencies below about 3 kHz compared to the Greenwood map. Elongating VTL causes the formant peaks to shift toward lower frequencies compared to shortening VTL, hence, the peaks fall in the region where there is no sufficient spectral resolution to resolve spectral shifts along the lower frequencies.

Overall, these results indicate that the large difference in overall mean JNDs ( $\delta = 3.35 \text{ st}$ ) between the linear and Greenwood partitioning maps for the ideal case simulated in this experiment supports the idea that an optimal frequency partitioning map may, in fact, help improve VTL sensitivity. Since there were only two maps in this experiment, in experiment 3, the Greenwood map was compared to two clinical maps to check whether it would also outperform the mapping available in standard clinical settings.

Moreover, experiment 3 attempts to remedy some of the limitations of experiments 1 and 2 by using more realistic simulations of electrode positions and filter partitioning according to some clinical frequency maps.

## V. EXPERIMENT 3: EFFECT OF FREQUENCY MISMATCH AND BAND PARTITIONING ON VTL SENSITIVITY

### A. Rationale

Experiments 1 and 2 revealed a significant effect of frequency mismatch and band partitioning on VTL JNDs, respectively. The data showed that the larger the mismatch, the worse the sensitivity to VTL differences became. Moreover, the fewer the channels allocated to the lower half of the frequency partition, the worse the VTL JNDs were.

The aim of this third experiment was to test the combined effect of frequency mismatch and band partitioning on VTL JNDs since this is a more realistic scenario in actual implants. The hypothesis was that a partitioning map with sufficient spectral resolution may still help preserve VTL-related cues, even under extreme frequency mismatch conditions. If this is the case, then it should manifest as a lack of interaction between the frequency partitioning and the mismatch. To test this, analysis filters were partitioned according to the linear and Greenwood maps used in experiment 2. In addition, to compare the Greenwood map's performance to that of clinical maps, the analysis filters were also partitioned according to the Cochlear and HiRes maps, as defined in Sec. II (General Methods; see panel 3 of Fig. 2).

To mimic the frequency mismatch observed in actual implants, the synthesis filters were partitioned based on the dimensions of the HiFocus Helix electrode array. This created two mismatch scenarios: a *minimal shift* if the simulated electrode array is inserted until the proximal marker, and a *maximal shift* if the array is inserted until the distal marker.

## B. Methods

### 1. Participants

The same participants who took part in experiment 2 participated in this experiment using the same apparatus and procedure as in experiment 2. Additionally, hearing thresholds between 8 kHz and 16 kHz were also measured with special headphones (Koss R/80 headphones, Koss Corporation, Milwaukee, WI) that were calibrated to a clinical audiometer by EMID (Electro Medical Instruments BV Doesburg, Doesburg, NL). This was done to ensure that participants could hear stimuli components falling in the higher frequency bands resulting from the basal-ward shift in the synthesis filters for the maximal shift condition (see panel 2 in Fig. 2). Under that setting, the most basal filter band was defined between 12.8 and 14.4 kHz.

### 2. Procedure

In this experiment, 16 experimental conditions were administered: 2 VTL manipulation types (elongating or shortening VTL)  $\times$  4 maps (analysis filter settings)  $\times$  2 frequency shift conditions (synthesis filter settings). In the training phase, the two VTL manipulation types were tested using both frequency shift conditions for only the Greenwood map (2 VTL manipulations  $\times$  1 map  $\times$  2 shift conditions = 4 conditions) to familiarize the participants with the procedure.

In addition, at the beginning of each run, a short preview block was provided to familiarize the participants with the VTL manipulation and band partitioning tested in this run. This was done because, based on a pilot experiment, it was observed that participants found this particular experiment too difficult due to the large number of different vocoders that forced them to readjust their strategy constantly. These preview blocks consisted of five words randomly chosen from the NVA corpus. Each word was vocoded using the parameters of the current condition and presented twice on the screen to the participant: once shown in blue to denote the reference VTL voice, and once again in red to indicate the target VTL voice. The participants were asked to listen to the difference between the red and blue versions of each word before the three-alternative forced choice task (3AFC) task began.

## C. Results and discussion

The mean JND distribution across participants for each analysis filter partitioning map is shown in Fig. 6, for minimal versus maximal shift conditions (left panel), and for elongating versus shortening VTL relative to the reference female voice (right panel).

A three-way repeated measures ANOVA was applied on the log-transformed VTL JNDs with analysis filter partitioning, frequency shift, and VTL manipulation type (elongating

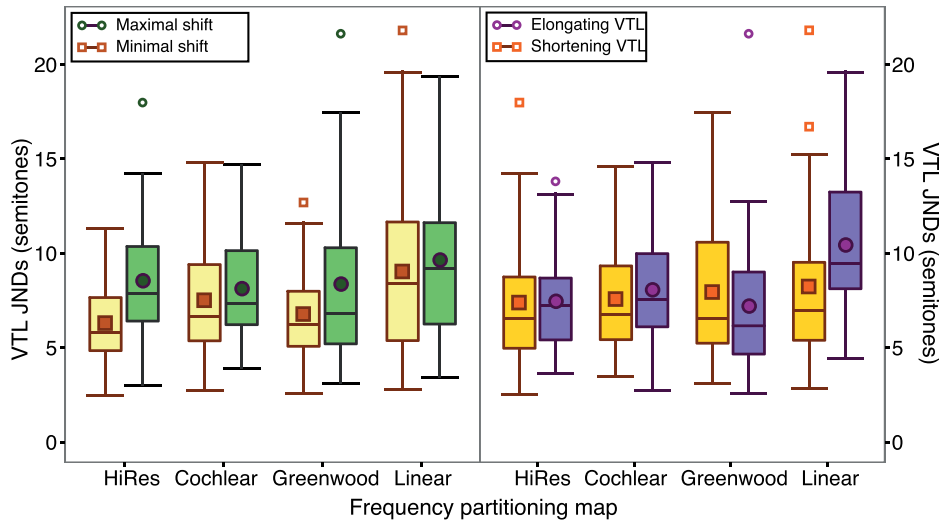


FIG. 6. (Color online) VTL JNDs shown as a function of analysis filter partitioning maps for minimal versus maximal shift (left), and for elongating versus shortening VTL relative to the reference female voice (right). The boxes extend from the lower to the upper quartile, and the middle line shows the median. The filled symbols (circle and square) show the means for maximal and minimal shift conditions, respectively (left), and for elongating and shortening VTL, respectively (right). The details of the boxplot are as described in Fig. 3.

or shortening) as repeated factors. Consistent with what was found in experiment 1, this analysis revealed a significant, albeit small, effect of frequency shift [ $F(1,14) = 21.45$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.038$ ], such that minimal shift yielded better (smaller) JNDs ( $\mu = 7.41$  st,  $\sigma = 3.49$  st) compared to the maximal shift condition ( $\mu = 8.67$  st,  $\sigma = 3.81$  st), irrespective of the analysis filter partitioning map.

In addition, the ANOVA showed a significant effect of frequency partitioning on VTL JNDs [ $F(3,52) = 19.13$ ,  $p < 0.01$ ,  $\eta_G^2 = 0.041$ ], which is in line with what was found in experiment 2, but again with a small effect size.

Only the interaction between the analysis filter partitioning and the VTL manipulation type was found to have a significant effect on VTL thresholds [ $F(3,42) = 6.81$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.025$ ]. This means that some partitioning maps better relay shorter VTLs compared to longer VTLs, while others do not.

No other interaction between the factors was found to significantly affect VTL JNDs: consistent with the proposed hypothesis, the interaction between analysis filter partitioning and frequency shift was not found to be significant [ $F(3,42) = 1.104$ ,  $p = 0.358$ ,  $\eta_G^2 = 0.007$ ]. This means that when sufficient spectral resolution is provided by the band partitioning map, VTL-related cues can still be sufficiently transmitted, even under extreme frequency mismatch conditions.

Pairwise  $t$ -tests with FDR correction revealed that only the linear map was significantly worse than the HiRes and Greenwood maps [linear versus HiRes:  $t(14) = 3.61$ ,  $p_{FDR} = 0.015$ ,  $\delta = 1.74$  st; linear versus Greenwood:  $t(14) = 3.55$ ,  $p_{FDR} = 0.015$ ,  $\delta = 1.58$  st], while there was no difference in VTL JNDs between the HiRes, Cochlear, and Greenwood maps, and the linear versus Cochlear maps ( $p_{FDR} > 0.18$  for all comparisons). This suggests that the resolution of the low-frequency components, where formants are defined, is important for the perception of VTL differences, and the clinical maps are not significantly worse than the Greenwood map, at least in simulation.

What is notable is how the different frequency partitioning maps compare to each other when VTL is elongated or shortened relative to the reference voice, as was observed in experiment 2. In the case where VTL was shortened with

respect to the reference voice, all four maps appeared to yield similar performance ( $p_{FDR} > 0.45$  for all pairwise comparisons under this condition). However, when VTL was elongated relative to the reference, the linear map yielded significantly worse (larger) JNDs compared to all other maps [linear versus HiRes:  $t(14) = 4.37$ ,  $p_{FDR} = 0.006$ ,  $\delta = 2.85$  st; linear versus Cochlear:  $t(14) = 2.84$ ,  $p_{FDR} = 0.047$ ,  $\delta = 2.32$  st; linear versus Greenwood:  $t(14) = 5.6$ ,  $p_{FDR} = 0.001$ ,  $\delta = 3.17$  st], while there was no difference in performance for all other maps under this condition ( $p_{FDR} > 0.14$ ). This means that increasing the resolution of the frequency partitioning map for frequencies below about 3 kHz is important for conveying different types of voices. In addition, the clinical maps tested in this experiment appear to convey such voice differences at least as well as the Greenwood map. It is only when the spectral resolution near the lower frequencies becomes sufficiently low, as is the case with the linear map, that transmission of these voice differences becomes compromised.

This behaviour can be explained by looking at the spectra of sounds from the output of each frequency map setup, as shown in Fig. 7. In the top panel, the spectral envelope of an unvoiced long vowel /a:/ is shown for three different VTL settings. The black solid line represents the vowel /a:/ of the reference speaker. The dotted red and dashed blue lines represent a VTL shift of  $-6$  st (shortening VTL, increasing formant frequency) and  $+6$  st (elongating VTL, decreasing formant frequency), respectively, as was done in Fig. 4. In the bottom panel, the spectral envelopes of the vowel are plotted against the synthesis filter frequencies under the minimal shift condition. The green arrows indicate the relative distance between the reference vowel and the VTL-shifted versions for all map conditions in the region around 3 kHz, where most formants are expected to lie. The larger this distance is between the reference and VTL-shifted versions, the easier it should be to differentiate the reference signal from the VTL-shifted one. This distance is much larger for the HiRes, Cochlear, and Greenwood maps compared to the linear map. In the case of the signals examined in Fig. 7, the  $\pm 6$  st difference in the unvoiced vowel translates to a difference between roughly 3.53 st and 4.74 st when the HiRes, Cochlear, or Greenwood maps are used as

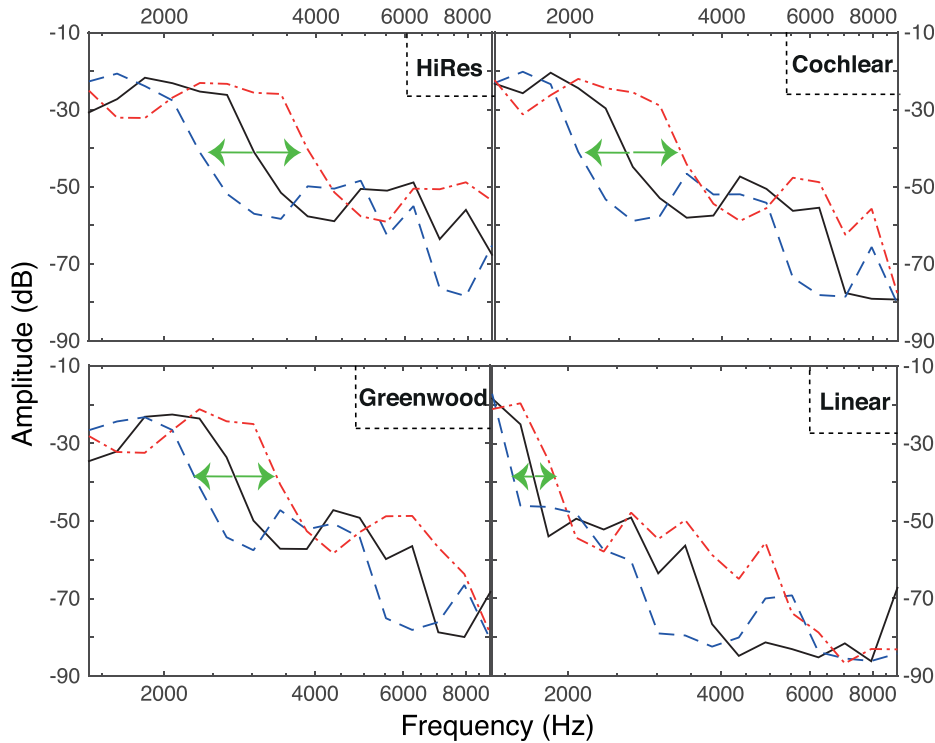
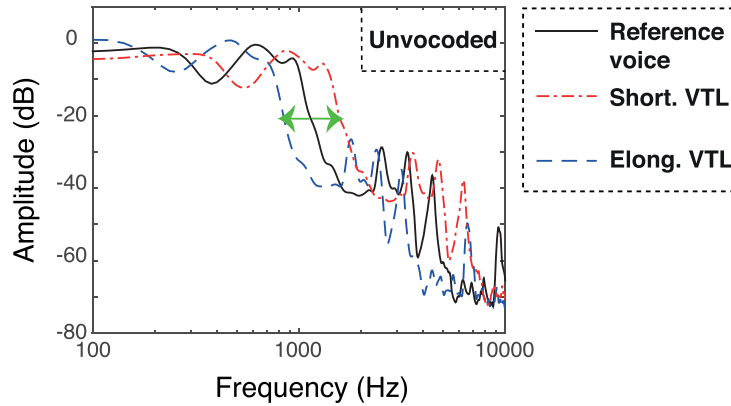


FIG. 7. (Color online) Spectral envelopes for long vowel /a/. The solid black line indicates the envelope of the vowel with the reference VTL. The dotted red and dashed blue lines indicate a VTL shift of  $-6$  st (shortening VTL) and  $+6$  st (elongating VTL), respectively. (Top) Spectra for the VTL-shifted vowel for the unvocoded case. (Middle, bottom) Spectra obtained from the output of the analysis filters and plotted versus the frequencies of the synthesis filters for the minimal shift condition. Green arrows indicate the relative distance between the VTL-shifted vowel and the reference version.

analysis filters. However, this  $\pm 6$  st difference is only translated to about a 2.95-st difference if the linear map is applied. These differences were computed as the mean of the semitone difference between the frequencies of the first three peaks in the reference signal, and the corresponding peaks in the VTL-shifted signals. Such an effect may be due to the inherently larger number of bands (12–13 bands) assigned to frequencies below about 3.5 kHz (a higher spectral resolution at those frequencies) for the HiRes, Cochlear, and Greenwood maps compared to the seven bands assigned to those frequencies under the linear map. This may explain the significantly larger JNDs observed for the linear map.

As for VTL JNDs being worse for elongating versus shortening VTL for the linear map, this can be explained by comparing the envelopes produced by the linear map to their unvocoded counterpart. Notice how the shapes of the spectral envelopes in the unvocoded version are somewhat maintained after applying the linear map to the reference voice (black solid line) and to its shortened VTL version (dotted red line). However, when VTL is elongated (dashed blue line), the shape of the spectral envelope is distorted after

applying the linear mapping. One might argue that the shape of the envelope is also somewhat distorted for the other three maps, however, the effect of having a larger distance between the VTL-shifted versions and the reference vowel compared to the linear map may provide more salient cues for the detection of VTL differences.

## VI. GENERAL DISCUSSION

In this study, the effect of frequency shift and band partitioning on VTL sensitivity were investigated both in isolation (experiments 1 and 2, respectively) and in conjunction (experiment 3). Results from all three experiments showed a dependency of VTL sensitivity on frequency mismatch (shift), filter slope (simulated channel interaction), and frequency band partitioning (spectral resolution near the lower frequencies), in addition to the interaction between the frequency partitioning and VTL manipulation.

Frequency mismatch, implemented as an increasing shift between the analysis and synthesis filters, worsened the sensitivity to VTL. Since formant cues are important for

both VTL perception, as well as for vowel identification, a frequency mismatch that affects VTL cues would also be expected to affect vowel identification. Indeed, the findings presented here are consistent with previous vocoder studies that reported a decline in vowel recognition scores as a function of increased frequency shift (Başkent and Shannon, 2004; Fitzgerald *et al.*, 2013; Fu and Shannon, 1999b; Shannon *et al.*, 1998).

Shallower filter slopes, simulating channel interaction, decreased the sensitivity to VTL differences. This is in agreement with the results reported by Gaudrain and Başkent (2015) for VTL sensitivity, and with those reported by Fu and Shannon (2002) and Shannon *et al.* (1998) for vowel recognition scores.

Band partitioning, simulated by decreasing the spectral resolution for frequencies below about 3 kHz (where the first three formants are usually represented) led to a reduction in sensitivity to VTL cues. This is consistent with the effect of band partitioning on vowel recognition scores reported in the literature (Fu and Shannon, 2002; McKay and Henshall, 2002; Shannon *et al.*, 1998). In the current study, the spectral resolution in the lower frequency region seems essential in conveying longer VTLs as efficiently as shorter VTLs. For example, all maps from experiment 3, except for the linear map, yielded similar performance for longer and shorter VTLs. The linear map hindered access to cues from longer VTLs more than for shorter VTLs. This means that if a map has no sufficient spectral resolution in the lower half of its frequency range, then differences between longer and shorter VTLs would not be sufficiently conveyed. In this study, since the reference VTL was that of a female, and transmission of longer VTL cues was impaired, this indicates that gender-related differences in voice cues carried by VTL may be compromised in such situations. Finally, because the effect of band partitioning was independent from that of frequency mismatch, a band partitioning map with sufficient spectral resolution may help mitigate some of the negative effects of mismatch on VTL sensitivity.

It is worth noting that the effects observed here, while statistically significant, had a small effect size and were obtained using only simulations of CI signal processing. Nonetheless, since band partitioning was found to improve VTL sensitivity despite the severity of the mismatch, it may be worthwhile to investigate the effect of band partitioning in CI users.

## VII. CONCLUSION

CI users exhibit poor perception of vocal cues, especially VTL, which may be a result of two effects. The first is the frequency mismatch between the frequencies received by the implant and those corresponding to the actual place of stimulation in the cochlea. The second is the poor spectral resolution in the implant arising from suboptimal frequency-to-electrode allocation mapping, which is seldom adjusted for each individual CI user. In this study, VTL JNDs were investigated as a function of frequency mismatch and band partitioning in vocoder simulations with NH listeners. Frequency mismatch was implemented as a shift between the vocoder

analysis and synthesis filters, while frequency band partitioning was applied to the analysis filters. VTL JNDs were found to depend on (1) the degree of mismatch and channel interaction between analysis and synthesis filters, (2) the analysis filter band partitioning, and (3) the interplay between the analysis filter partitioning and the VTL manipulation type. In particular, sufficient resolution near the low frequencies of the frequency band partitioning map was found to improve VTL JNDs, irrespective of the degree of frequency mismatch. Thus, this effect of band partitioning may be worthwhile to investigate in CI listeners, since it may likely affect their VTL discrimination as well, and especially that it does not require modifications to actual device design.

## ACKNOWLEDGMENTS

The work presented here was jointly funded by Advanced Bionics (AB), the University Medical Center Groningen (UMCG), and the PPP-subsidy of the Top consortia for Knowledge and Innovation of the Dutch Ministry of Economic Affairs. The authors have further been supported by a Rosalind Franklin Fellowship from the University Medical Center Groningen, University of Groningen, and the VICI Grant No. 016.VICI.170.111 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw), and funds from the Heinsius Houbolt Foundation. This work was conducted in the framework of the LabEx CeLyA (“Centre Lyonnais d’Acoustique”, ANR-10-LABX-0060/ANR-11-IDEX-0007) operated by the French National Research Agency, and is also part of the research program of the Otorhinolaryngology Department of the University Medical Center Groningen: Healthy Aging and Communication. The authors would like to thank Bert Maat, Emile de Klein, Sander Ubbink, and Jeanne Clarke for their help with audiometry measurements, Frits Leemhuis for his assistance during the audiometer calibrations, and Paolo Toffanin and Enja Jung for their help with stimulus calibration. The authors would also like to thank all colleagues who helped pilot this study, all the participants who volunteered, and all the staff of the Keel-, Neus-, en Oorheelkunde (KNO) clinic at the University Medical Center Groningen (UMCG).

## APPENDIX: FREQUENCY BAND PARTITIONING MAPS IN THE LITERATURE

Some of the frequency band partitioning maps proposed in the literature were replotted in Fig. 8. This was done to help the reader compare the different maps used in the literature because different studies used different representations (equations or different types of figures).

Only a selected number of the frequency partitioning maps described in those studies are shown to aid in visual comparison with the ones chosen for this study [Fig. 8(H)]. Figure 8(A) shows the three maps used in the study by Shannon *et al.* (1998). In that study, a linear and a Greenwood map (Greenwood, 1990) were tested, along with an intermediate map between those two extremes. In Figure 8(B), only four of the ten maps used by Fu and Shannon (1999b) are

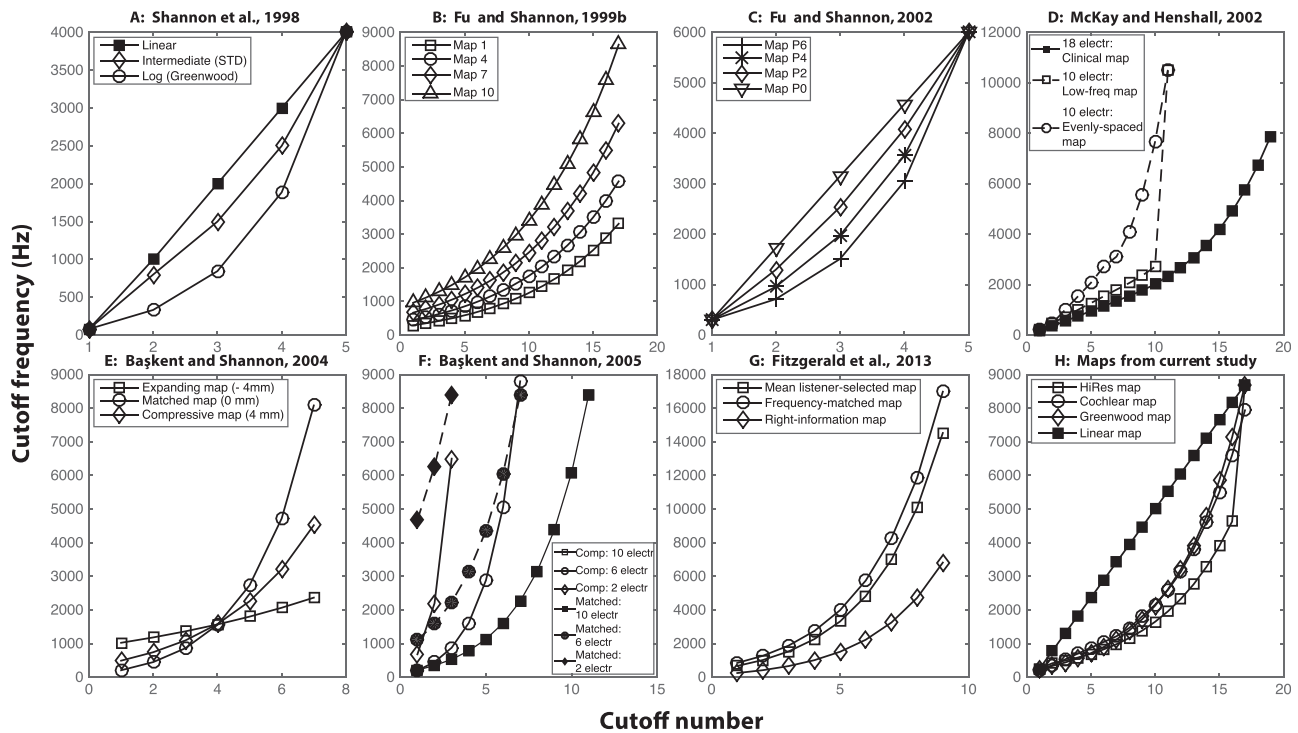


FIG. 8. Different frequency partitioning maps specified in the literature compared to the four maps presented in this study. (A) The linear and logarithmic (Greenwood) partitioning maps used by Shannon *et al.* (1998). The standard condition (STD) map is an intermediate map between both the linear and log ones. (B) Four of the ten maps used by Fu and Shannon (1999b), all partitioned according to the Greenwood formula [see Eq. (1)]. (C) Parametric map manipulations from linear to logarithmic as defined by Fu and Shannon (2002). (D) The maps defined by McKay and Henshall (2002). Filled symbols represent 18-electrode maps, while hollow symbols indicate 10-electrode maps. (E) The *expanding*, *matched*, and *compressive* maps described by Başkent and Shannon (2004). Only the most extreme manipulations are provided here. (F) *Compressed* (hollow symbols) and *matched* maps (filled symbols) defined by Başkent and Shannon (2005). (G) The three maps used by Fitzgerald *et al.* (2013) in phoneme and word recognition tasks. (H) Description of the four maps used in this study.

depicted. This is because, in that study, the authors defined ten maps that were partitioned according to the Greenwood formula but were systematically shifted away toward more basal frequencies relative to map 1. Figure 8(C) depicts only four of the six maps defined by Fu and Shannon (2002), which varied systematically from a purely linear partitioning (map P0) to a purely logarithmic one (map P6). Figure 8(D) shows only three maps from the ones introduced by McKay and Henshall (2002). The first seven channels of the *evenly spaced map* are almost linearly partitioned compared to both the *clinical* and *low-frequency* maps. The *low-frequency map* (empty squares with dashed lines) assigns nine out of the ten channels to low frequencies below 3 kHz, while the last channel spans a large range of frequencies up to 10 kHz, hence the sharp rise in the function. Consequently, this partitioning has a higher resolution at the lower frequencies compared to the *evenly spaced map*. Figure 8(E) provides only the most extreme manipulations described by Başkent and Shannon (2004). Notice also how the partitioning varies from a linear function to a log-like function. Figure 8(F) shows the compressed and matched maps defined by Başkent and Shannon (2005). Figure 8(G) shows the analysis filter partitioning maps used by Fitzgerald *et al.* (2013). The *mean-listener-selected map* is the mean of all individual maps selected by the participants in a self-fitting procedure, the *frequency-matched map* is the map matching the synthesis filters of the vocoder used in their experiment to the analysis filters, and the *right-information map* is based on a standard clinical map. Notice that, on average, participants prefer the map with no

mismatch compared to the clinical map, in which the analysis filter partitioning was different than the synthesis filter partitioning. Finally, Figure 8(H) shows the analysis filter partitioning maps used in the current study.

Assmann, P. F., and Summerfield, Q. (1990). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies." *J. Acoust. Soc. Am.* **88**, 680–697.

Avci, E., Nauwelaers, T., Lenarz, T., Hamacher, V., and Kral, A. (2014). "Variations in microanatomy of the human cochlea." *J. Comp. Neurol.* **522**, 3245–3261.

Başkent, D., and Gaudrain, E. (2016). "Musician advantage for speech-on-speech perception." *J. Acoust. Soc. Am.* **139**, EL51–EL56.

Başkent, D., Gaudrain, E., Tamati, T. N., and Wagner, A. (2016). "Perception and psychoacoustics of speech in cochlear implant users," in *Scientific Foundations of Audiology: Perspectives from Physics, Biology, Modeling, and Medicine*, edited by A. T. Cacace, E. de Kleine, A. G. Holt, and P. van Dijk (Plural, San Diego), pp. 285–319.

Başkent, D., and Shannon, R. V. (2004). "Frequency-place compression and expansion in cochlear implant listeners." *J. Acoust. Soc. Am.* **116**, 3130–3140.

Başkent, D., and Shannon, R. V. (2005). "Interactions between cochlear implant electrode insertion depth and frequency-place mapping." *J. Acoust. Soc. Am.* **117**, 1405–1416.

Benjamini, Y., and Yekutieli, D. (2001). "The control of the false discovery rate in multiple testing under dependency." *Ann. Statist.* **29**, 1165–1188.

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. F., O'Leary, S., Rousset, A., Schauwers, K., Skarzynski, H., Skarzynski, P. H., Sterkers, O., Terranti, A., Truy, E., Van de Heyning, P., Venail, F., Vincent, C., and Lazard, D. S. (2012). "Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients." *Audiol. Neurotol.* **18**, 36–47.

- Boëx, C., de Balthasar, C., Kós, M.-I., and Pelizzone, M. (2003). "Electrical field interactions in different cochlear implant systems," *J. Acoust. Soc. Am.* **114**, 2049–2057.
- Bosman, A. J., and Smoorenburg, G. F. (1995). "Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment," *Audiology* **34**, 260–284.
- Brokx, J., and Nootboom, S. (1982). "Intonation and the perceptual separation of simultaneous voices," *J. Phonetics* **10**, 23–36.
- Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.* **109**, 1101–1109.
- Carlyon, R. P., and Shackleton, T. M. (1994). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?," *J. Acoust. Soc. Am.* **95**, 3541–3554.
- Chiba, T., and Kajiyama, M. (1941). *The Vowel: Its Nature and Structure* (Tokyo-Kaiseikan, Tokyo).
- Darwin, C. J., Brungart, D. S., and Simpson, B. D. (2003). "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," *J. Acoust. Soc. Am.* **114**, 2913–2922.
- De Balthasar, C., Boex, C., Cosendai, G., Valentini, G., Sigrist, A., and Pelizzone, M. (2003). "Channel interactions with high-rate biphasic electrical stimulation in cochlear implant subjects," *Hear. Res.* **182**, 77–87.
- Di Nardo, W., Scorpecci, A., Giannantonio, S., Cianfrone, F., and Paludetti, G. (2011). "Improving melody recognition in cochlear implant recipients through individualized frequency map fitting," *Eur. Arch. Otorhinolaryngol.* **268**, 27–39.
- Dowell, R. C., Mecklenburg, D. J., and Clark, G. M. (1986). "Speech recognition for 40 patients receiving multichannel cochlear implants," *Arch. Otolaryngol. Head. Neck. Surg.* **112**, 1054–1059.
- Drullman, R., and Bronkhorst, A. W. (2004). "Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers," *J. Acoust. Soc. Am.* **116**, 3090–3098.
- Dudley, H. (1939). "The vocoder," *Bell Labs Rec.* **18**, 122–126.
- Evans, S., Neave, N., and Wakelin, D. (2006). "Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice," *Biol. Psychol.* **72**, 160–163.
- Fant, G. (1960). *Acoustic Theory of Speech Perception* (Mouton, The Hague).
- Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.
- Finley, C. C., Holden, T. A., Holden, L. K., Whiting, B. R., Chole, R. A., Neely, G. J., Hullar, T. E., and Skinner, M. W. (2008). "Role of electrode placement as a contributor to variability in cochlear implant outcomes," *Otol. Neurotol.* **29**, 920–928.
- Fitch, W. T., and Giedd, J. (1999). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," *J. Acoust. Soc. Am.* **106**, 1511–1522.
- Fitzgerald, M. B., Sagi, E., Morbiwala, T. A., Tan, C.-T., and Svirsky, M. A. (2013). "Feasibility of real-time selection of frequency tables in an acoustic simulation of a cochlear implant," *Ear Hear.* **34**, 763–772.
- Franke-Trieger, A., and Mürbe, D. (2015). "Estimation of insertion depth angle based on cochlea diameter and linear insertion depth: A prediction tool for the CI422," *Eur. Arch. Otorhinolaryngol.* **272**, 3193–3199.
- Friesen, L. M., Shannon, R. V., Başkent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.
- Fu, Q.-J., and Shannon, R. V. (1999a). "Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus-22 cochlear implant," *Ear Hear.* **20**, 332–344.
- Fu, Q.-J., and Shannon, R. V. (1999b). "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," *J. Acoust. Soc. Am.* **105**, 1889–1900.
- Fu, Q.-J., and Shannon, R. V. (2002). "Frequency mapping in cochlear implants," *Ear and Hearing* **23**, 339–348.
- Fu, Q.-J., Shannon, R. V., and Wang, X. (1998). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," *J. Acoust. Soc. Am.* **104**, 3586–3596.
- Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., and Başkent, D. (2014). "Gender categorization is abnormal in cochlear implant users," *J. Assoc. Res. Otolaryngol.* **15**, 1037–1048.
- Gaudrain, E., and Başkent, D. (2015). "Factors limiting vocal-tract length discrimination in cochlear implant simulations," *J. Acoust. Soc. Am.* **137**, 1298–1308.
- Gaudrain, E., and Başkent, D. (2018). "Discrimination of voice pitch and vocal-tract length in cochlear implant users," *Ear Hear.* **39**, 226–237.
- Grasmeder, M. L., Verschuur, C. A., and Batty, V. B. (2014). "Optimizing frequency-to-electrode allocation for individual cochlear implant users," *J. Acoust. Soc. Am.* **136**, 3313–3324.
- Greenwood, D. D. (1990). "A cochlear frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.* **87**, 2592–2605.
- Hanekom, J. J., and Shannon, R. V. (1998). "Gap detection as a measure of electrode interaction in cochlear implants," *J. Acoust. Soc. Am.* **104**, 2372–2384.
- Henry, B. A., and Turner, C. W. (2003). "The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 2861–2873.
- Hillenbrand, J. M., and Clark, M. J. (2009). "The role of f (0) and formant frequencies in distinguishing the voices of men and women," *Atten. Percept. Psychophys.* **71**, 1150–1166.
- Ives, D. T., Smith, D. R. R., and Patterson, R. D. (2005). "Discrimination of speaker size from syllable phrases," *J. Acoust. Soc. Am.* **118**, 3816–3822.
- Kawahara, H., and Irino, T. (2005). "Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Springer, Boston, MA), pp. 167–180.
- Landsberger, D. M., Svrakic, M., Roland, J. T., and Svirsky, M. (2015). "The relationship between insertion angles, default frequency allocations, and spiral ganglion place pitch in cochlear implants," *Ear Hear.* **36**, e207–e213.
- Lee, J. H., and Humes, L. E. (2012). "Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background," *J. Acoust. Soc. Am.* **132**, 1700–1717.
- Leigh, J. R., Henshall, K. R., and McKay, C. M. (2004). "Optimizing frequency-to-electrode allocation in cochlear implants," *J. Am. Acad. Audiol.* **15**, 574–584.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Licklider, J. (1954). "'Periodicity' pitch and 'place' pitch," *J. Acoust. Soc. Am.* **26**, 945–945.
- Lieberman, P., and Blumstein, S. E. (1988). "Source-filter theory of speech production," in *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge University Press, Cambridge, UK), pp. 34–50.
- Massida, Z., Marx, M., Belin, P., James, C., Fraysse, B., Barone, P., and Deguine, O. (2013). "Gender categorization in cochlear implant users," *J. Speech. Lang. Hear. Res.* **56**, 1389–1401.
- McKay, C. M., and Henshall, K. R. (2002). "Frequency-to-electrode allocation and speech perception with cochlear implants," *J. Acoust. Soc. Am.* **111**, 1036–1044.
- Meister, H., Fürsen, K., Streicher, B., Lang-Roth, R., and Walger, M. (2016). "The use of voice cues for speaker gender recognition in cochlear implant recipients," *J. Speech. Lang. Hear. Res.* **59**, 546–556.
- Müller, J. (1848). *The Physiology of the Senses, Voice, and Muscular Motion, with the Mental Faculties*, Taylor, Walton & Maberly.
- Omran, S. A., Lai, W., Büchler, M., and Dillier, N. (2011). "Semitone frequency mapping to improve music representation for nucleus cochlear implants," *EURASIP J. Audio, Speech, Music Process.* **2011**, 2.
- Oxenham, A. J. (2008). "Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants," *Trends Amplif.* **12**, 316–331.
- Patterson, R. D., and Smith, D. R. (2003). *The Perception of Scale in Vowels* (Cambridge Univ. Press, Cambridge, UK).
- Pyschny, V., Landwehr, M., Hahn, M., Walger, M., von Wedel, H., and Meister, H. (2011). "Bimodal hearing and speech perception with a competing talker," *J. Speech. Lang. Hear. Res.* **54**, 1400–1415.
- Qin, M. K., and Oxenham, A. J. (2003). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," *J. Acoust. Soc. Am.* **114**, 446–454.
- R Core Team (2014). "R: A language and environment for statistical computing," version 3.1.2, R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/> (Last viewed 4 June 2018).
- Shannon, R. V. (1983). "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," *Hear. Res.* **12**, 1–16.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.



- Shannon, R. V., Zeng, F.-G., and Wygonski, J. (1998). "Speech recognition with altered spectral distribution of envelope cues," *J. Acoust. Soc. Am.* **104**, 2467–2476.
- Skinner, M. W., Holden, T. A., Whiting, B. R., Voie, A. H., Brunsten, B., Neely, J. G., Saxon, E. A., Hullar, T. E., and Finley, C. C. (2007). "*In vivo* estimates of the position of advanced bionics electrode arrays in the human cochlea," *Ann. Otol. Rhinol. Laryngol.* **116**, 2–24.
- Skuk, V. G., and Schweinberger, S. R. (2014). "Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender," *J. Speech. Lang. Hear. Res.* **57**, 285–296.
- Smith, D. R. R., and Patterson, R. D. (2005). "The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age," *J. Acoust. Soc. Am.* **118**, 3177–3186.
- Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (2005). "The processing and perception of size information in speech sounds," *J. Acoust. Soc. Am.* **117**, 305–318.
- Stevens, K. N., and House, A. S. (1955). "Development of a quantitative description of vowel articulation," *J. Acoust. Soc. Am.* **27**, 484–493.
- Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F.-G. (2007). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," *J. Acoust. Soc. Am.* **122**, 1069–1078.
- Stickney, G. S., Zeng, F.-G., Litovsky, R., and Assmann, P. (2004). "Cochlear implant speech recognition with speech maskers," *J. Acoust. Soc. Am.* **116**, 1081–1091.
- Sylmar, C. (2005). "HiRes 90K surgeon's manual for the HiFocus Helix and HiFocus Ij electrodes (Rev E)," Advanced Bionics Corporation. Available at <https://advancedbionics.com/content/dam/advancedbionics/Documents/libraries/Audiologists-&-Surgeons-Library/Surgical-Manuals-&-Catalogs/hires-90k-surgeons-manual-hifocus-helix-1j.pdf> (Last viewed 4 June 2018).
- Tan, C.-T., Martin, B., and Svirsky, M. A. (2017). "Pitch matching between electrical stimulation of a cochlear implant and acoustic stimuli presented to a contralateral ear with residual hearing," *J. Am. Acad. Audiol.* **28**, 187–199.
- Townshend, B., and White, R. L. (1987). "Reduction of electrical interaction in auditory prostheses," *IEEE Trans. Biomed. Eng. BME*-**34**, 891–897.
- Tyler, R. S., Abbas, P., Tye-Murray, N., Gantz, B. J., Knutson, J. F., McCabe, B. F., Lansing, C., Brown, C., Woodworth, G., Hinrichs, J., and Kuk, F. (1988). "Evaluation of five different cochlear implant designs: Audiologic assessment and predictors of performance," *Laryngoscope* **98**, 1100–1106.
- van der Marel, K. S., Briaire, J. J., Wolterbeek, R., Snel-Bongers, J., Verbist, B. M., and Frijns, J. H. (2014). "Diversity in cochlear morphology and its influence on cochlear implant electrode position," *Ear Hear.* **35**, e9–e20.
- Venail, F., Mathiolon, C., Menjot de Champfleur, S., Piron, J. P., Sicard, M., Villemus, F., Vessigaud, M. A., Sterkers-Artieres, F., Mondain, M., and Uziel, A. (2015). "Effects of electrode array length on frequency-place mismatch and speech perception with cochlear implants," *Audiol. Neurootol.* **20**, 102–111.
- Vestergaard, M. D., Fyson, N. R., and Patterson, R. D. (2009). "The interaction of vocal characteristics and audibility in the recognition of concurrent syllables," *J. Acoust. Soc. Am.* **125**, 1114–1124.
- Winn, M. B., Won, J. H., and Moon, I. J. (2016). "Assessment of spectral and temporal resolution in cochlear implant users using psychoacoustic discrimination and speech cue categorization," *Ear Hear.* **37**, e377–e390.