Nonlinear frequency compression: Influence of start frequency and input bandwidth on consonant and vowel recognition^{a)}

Joshua M. Alexander^{b)}

Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana 47907, USA

(Received 11 April 2015; revised 26 January 2016; accepted 1 February 2016; published online 23 February 2016)

By varying parameters that control nonlinear frequency compression (NFC), this study examined how different ways of compressing inaudible mid- and/or high-frequency information at lower frequencies influences perception of consonants and vowels. Twenty-eight listeners with mild to moderately severe hearing loss identified consonants and vowels from nonsense syllables in noise following amplification via a hearing aid simulator. Low-pass filtering and the selection of NFC parameters fixed the output bandwidth at a frequency representing a moderately severe (3.3 kHz, group MS) or a mild-to-moderate (5.0 kHz, group MM) high-frequency loss. For each group ($n = 14$), effects of six combinations of NFC start frequency (SF) and input bandwidth [by varying the compression ratio (CR)] were examined. For both groups, the 1.6 kHz SF significantly reduced vowel and consonant recognition, especially as CR increased; whereas, recognition was generally unaffected if SF increased at the expense of a higher CR. Vowel recognition detriments for group MS were moderately correlated with the size of the second formant frequency shift following NFC. For both groups, significant improvement (33%–50%) with NFC was confined to final /s/ and /z/ and to some VCV tokens, perhaps because of listeners' limited exposure to each setting. No set of parameters simultaneously maximized recognition across all tokens.

© 2016 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4941916>]

[FJG]

Pages: 938–957

I. INTRODUCTION

A. Frequency-lowering amplification and nonlinear frequency compression (NFC)

Hearing aid users often have limited access to important high-frequency speech information (Stelmachowicz *et al.*, 2001; Stelmachowicz *et al.*, 2002; Stelmachowicz *et al.*, 2004; Wolfe *et al.*, 2011). For listeners with mild to moderate sensorineural hearing loss (SNHL), this can occur because hearing aids are unable to provide sufficient high-frequency amplification or cannot do so without feedback or without causing unacceptable sound quality and/or excessive loudness (Moore *et al.*, 2008). For listeners with more severe SNHL, the inner hair cells that code these frequencies may be absent or non-functioning, possibly rendering amplification in this region less useful (e.g., Baer *et al.*, 2002; Vickers *et al.*, 2001). Frequency-lowering techniques are a means of re-introducing high-frequency speech cues for listeners with SNHL (Simpson 2009; Alexander *et al.*, 2014; Brennan *et al.*, 2014; Picou *et al.*, 2015; Wolfe *et al.*, 2015). The premise behind all frequency-lowering techniques is to use all or part of the audible low-frequency spectrum to code parts of the inaudible high-frequency spectrum. Most often, this entails “moving” part of the low-frequency spectrum that would otherwise be aided normally, in order to make room

for the additional high-frequency information (i.e., compression techniques) or superimposing the recoded high-frequency information on top of the existing low-frequency spectrum (i.e., transposition techniques). See Simpson (2009) and Alexander (2013) for a review of past and present frequency-lowering techniques and associated research findings.

As discussed by Alexander (2013), potential benefit associated with re-introducing high-frequency speech cues into the audible frequency range also comes with increased risk of distorting low-frequency speech cues. This tradeoff may be responsible in part for the varied outcomes observed within and between studies investigating frequency-lowering amplification (McCreery *et al.*, 2012; Alexander, 2013), especially a class of techniques known as “nonlinear frequency compression (NFC),” which is the focus of this paper. With NFC, the spectrum below a programmable start frequency (SF) is unaltered. Therefore, the SF can directly influence the fidelity of the low-frequency speech signal. The high-frequency spectrum is compressed toward the SF by an amount determined by the compression ratio (CR). Frequency reassignment occurs on a log scale with the technique implemented by Phonak (see Simpson *et al.*, 2005), which introduced the first commercial hearing aids with NFC, although alternative frequency remapping functions using NFC exist (e.g., “spectral envelope decimation” in Alexander *et al.*, 2014). As verified by Alexander (2013), frequency compression using the Simpson *et al.* method results in a logarithmic reduction in bandwidth (BW) and subsequent reduction on a psychophysical frequency scale.

^{a)}Portions of this work were presented at the 39th Annual Scientific and Technology Conference of the American Auditory Society, Scottsdale, Arizona.

^{b)}Electronic mail: alexan14@purdue.edu

For example, when $CR = 2.0:1$ (or, simply $CR = 2.0$, implying unity in the denominator) an input frequency distance that spans two auditory filters in the un-impaired ear will span only one auditory filter in the lowered output. In this way, the CR might influence the usability of the high-frequency information after lowering, with smaller CRs being more favorable for maintaining spectral features such as gross spectral shape. However, as discussed below, CR is dependent on other critical parameters, including SF, that together determine how frequency is remapped with NFC.

Because the SF and CR both influence how frequencies are remapped, there is a large combination of ways the un-aidable high-frequency spectrum can be compressed into the audible range of the listener. It is not clear how these parameters should be set for individual losses or whether the best settings depend on the sound class being lowered (e.g., consonants vs vowels or comparisons between subclasses of phonemes within the consonants/vowels). Figure 1 illustrates the tradeoffs a clinician encounters when using NFC to remap speech into a fixed output BW. Input frequency is represented along the abscissa and the frequencies where they are moved to in the output are represented along the ordinate. In each panel, the solid black line represents the frequency input-output function, with the black dot corresponding to the SF where NFC begins. The dotted horizontal gray line represents a hypothetical output BW, above which the aided signal is inaudible, and the intersecting vertical dotted line corresponds to the input BW (the maximum input frequency represented in the audible range after lowering). Panel (a) shows a setting with a low SF and a low CR. On the one hand, a low SF might be detrimental for phonemes that rely heavily on formant frequency, especially vowels. On the other hand, a low SF could be beneficial because it allows for a lower CR and better spectral resolution of the frequency-lowered speech for a fixed input BW [compare to panel (b)] and because it allows a greater amount of high-frequency information to be lowered if CR is fixed [compare to panel (c)]. Panel (b) shows the opposite case in which a higher SF is traded for a higher CR; that is, preservation of more low frequency information comes at the expense of reduced spectral resolution of the recoded high-frequency

information. Panel (c) shows a third option in which a high SF is combined with a low CR. While this option seems to mitigate the side effects of the other two options, it comes at the expense of a reduced input BW, meaning that less high-frequency information is represented in the frequency-lowered signal.

B. Literature review on the effects of NFC parameter selection on speech recognition

The selection of NFC parameters often varies across individuals between research studies and even within research studies, thereby making it difficult to make conclusions about their effects on speech recognition outcomes. To the author's knowledge, only the studies reviewed below have investigated the perceptual effects of varying NFC parameters in the same group of listeners and none have controlled for the various acoustic interdependencies between parameters.

1. Input bandwidth (BW)

Results from [McCreery et al. \(2013\)](#) and [Alexander et al. \(2014\)](#) provide evidence that fricative and affricate recognition improves as input BW increases. [McCreery et al. \(2013\)](#) used the same hearing simulator as described in this paper to implement wide dynamic range compression (WDRC) and NFC for nonsense words differing only in a fricative or affricate in the initial or final position. Normal-hearing adults identified the stimuli processed under three conditions for each of three simulated audiograms with output BWs of 2.5, 3.6, and 5.2 kHz. In addition to a low-pass control condition processed only with WDRC that simulated conventional hearing aid amplification, they tested two NFC settings that varied in the input BWs. The authors concluded that increasing the input BWs improved speech recognition. [Alexander et al. \(2014\)](#) also reported improved recognition as a function of input BW for listeners with SNHL who identified frequency-compressed fricatives and affricates in a vowel-consonant (VC) context with the vowel /i/ mixed with noise at a 10 dB signal-to-noise ratio (SNR) (these are the same as the "high-frequency stimuli" described in Sec. II).

FIG. 1. Schematic of hypothetical frequency remapping functions that demonstrate tradeoffs when using NFC to remap speech into a fixed output BW (horizontal dotted line). The abscissa and ordinate represent the input and output frequencies, respectively, and the solid black line in each panel represents the frequency input-output function, with the black dot corresponding to the start frequency. Where the input-output function intersects with the output BW (vertical dashed line) is the input BW, the maximum input frequency represented in the audible range after lowering. (a) Setting with a low SF and a low CR; (b) the opposite in which a higher SF is traded for a higher CR in order to maintain an equivalent input BW; (c) a high SF is combined with a low CR, thereby, sacrificing input BW.

2. Compression ratio (CR) and Start frequency (SF)

Ellis and Munro (2013) tested young, normal-hearing adults in two conditions with NFC and one condition without NFC. Both NFC conditions had a low SF, 1.6 kHz, and one had a CR = 2.0 and the other had a CR = 3.0. There was no attempt to control either input BW or output BW. The authors reported significant differences between all three conditions for recognition of sentences in noise, with ~80% recognition for sentences without NFC, ~40% for CR = 2.0, and ~30% for CR = 3.0. These results are consistent with the idea that implementing high CRs with low SFs can be detrimental to speech recognition.

Using adults and children with normal hearing and with SNHL, Parsa *et al.* (2013) examined the effects of NFC parameters on quality ratings of speech and music. CR varied from 2.0 to 10.0 and SF varied from 1.6 to 3.15 kHz. Even though there was no attempt to control either input BW or output BW, they found that listeners were influenced more by manipulations of SF than of CR, with sound quality ratings gradually decreasing as SF decreased below 3.0 kHz.

Souza *et al.* (2013) also conducted an investigation of the effects of SF and CR on sentence intelligibility across a wide range of SNRs in older adults with normal hearing or SNHL. Three levels of SF (1.0, 1.5, and 2.0 kHz) that were appropriate only for listeners with moderately severe SNHL, were crossed with three levels of CR (1.5, 2.0, and 3.0). (It should be noted that the form of NFC used by Souza *et al.* was different from what is clinically available and involved lowering only the most intense parts of the high-frequency spectrum with an unspecified frequency remapping function using a nominal CR that may or may not correspond to the same mathematical construct as other studies, including the current study.) In addition, the lowest SF tested was outside the range of clinically available options (≥ 1.5 kHz), there was no attempt to control either input or output BW, and only linear amplification with individualized frequency shaping was provided. Despite these differences from clinical implementations of NFC, it is noteworthy that the negative effects of NFC on sentence recognition increased as SF decreased and as CR increased and that there were minimal effects of NFC on recognition when SF = 2.0 kHz, regardless of CR, and when CR = 1.5, regardless of SF.

In summary, some of the studies reviewed here (McCreery *et al.*, 2013; Alexander *et al.*, 2014) suggest that across NFC conditions, parameter selections that generate greater effective input BW might be more favorable for speech recognition. However, this likely depends on output BW because when output BW is low, high input BW cannot be achieved without significantly increasing CR and/or decreasing SF, the combination of which (low SF with high CR) is expected to be detrimental to speech recognition based on previous findings (e.g., Ellis and Munro, 2013; Parsa *et al.*, 2013; Souza *et al.*, 2013).

C. Study rationale

Given the interdependency between input BW, CR, and SF, it is uncertain whether CR should be kept low to maintain spectral resolution or should be increased so that a

greater amount of high-frequency information can be lowered into the range of audibility and whether this effect depends on SF (e.g., lower CRs might be best for low SFs, but less important for high SFs where formant frequencies are less critical). The experiments described here answer these questions by testing consonant and vowel recognition with different combinations of SF and input BW, with the necessary CR derived to fit the selected values of these parameters. Input BW was manipulated instead of CR in order to control for the amount of information in the frequency-lowered signal. In addition, it was determined that this method resulted in a sufficient number of CRs within and between groups that were of similar range so that pertinent comparisons could be made.

Furthermore, nonsense monosyllables that varied only in one consonant or vowel were tested because it was hypothesized that tradeoffs affecting speech recognition would depend on the acoustic composition of the individual phonemes. That is, because the dominant cues for vowel identity rely on relationships between low- to mid-frequency formants (Ladefoged and Broadbent, 1957), which may be altered by NFC, no combination of settings was hypothesized to improve recognition. Therefore, vowel stimuli were used to assess the perceptual harm that might result from application of the technology. Specifically, it was hypothesized that detriments to vowel recognition would be proportional to the amount by which NFC shifted the relative formant frequencies, particularly F_2 , which systematically vary as a function of the vocal tract characteristics of the talker (higher for the shorter vocal tracts in children compared to women, and higher for women compared to men) and as a function of the individual phoneme.

With consonant stimuli, perceptual outcomes were less predictable because they contain a mixture of low- and high-frequency cues that may depend on the manner of articulation. That is, in a VCV context without frequency lowering, fricatives generally have more high-frequency information and require a greater bandwidth for maximum recognition than stops, followed by approximants (liquids and glides), while affricates and nasals in this context require very little high-frequency information to achieve maximum recognition (Alexander, 2010). Because of this difference in acoustic composition, the amount of benefit from NFC is hypothesized to be proportional to the bandwidth at which maximum recognition occurs for the un-lowered speech.

To demonstrate the differences in hypothesized benefit from NFC for vowels and consonants, Fig. 2 shows time waveforms and spectrograms for the vowel-consonant syllable /is/ after simulating a loss of audibility above 3.3 kHz. The insets in each panel show the spectrum of the steady portion of the vowel with the vertical line demarcating the original frequency of the second formant, F_2 . Compared with the original wideband signal in panel (a), the low-pass filtered signal in panel (b) shows that F_1 , F_2 , and part of F_3 associated with /i/ during the first 250 ms are preserved but there is essentially a total loss of high-frequency energy associated with the frication from the following /s/. Panel (c) shows the acoustic consequences in the time and frequency domains of an NFC setting with a 1.6 kHz SF and a 9.1 kHz

FIG. 2. Time waveforms and spectrograms for the vowel-consonant syllable /is/: (a) the wideband source signal; (b) low-pass filtered at 3.3 kHz to simulate a loss of audibility associated with moderately severe hearing loss; (c) processed with nonlinear frequency compression (NFC) using settings that reduce the range of frequencies from 1.6 to 9.1 kHz to 1.6 to 3.3 kHz. The inlets in each panel show the spectrum of the steady portion of the vowel with the dark vertical line demarcating the original frequency of the second formant and with units on the abscissa and ordinate representing 1 kHz and 10 dB, respectively.

input BW. The vowel portion shows the presence of the first three formants, but with altered frequency relationships: F_2 has been shifted by 0.74 kHz and F_3 has been shifted by 1.06 kHz. Vowel recognition might also be negatively affected by reducing the overall vowel space (Pols *et al.*, 1969; Klein *et al.*, 1970; Neel, 2008), thereby increasing confusion between frequency-lowered vowels sharing similar formant frequencies and by restricting the range of formant frequency movements (e.g., Hillenbrand and Nearey, 1999). Similarly, consonant recognition might be negatively affected by reducing the frequency extent of the formant transitions between consonants and vowels (e.g., Sussman *et al.*, 1991; Dorman and Loizou, 1996). Talker normalization and identification might also be negatively affected by changes in the formant frequencies (e.g., Bachorowski and Owren, 1999). On the other hand, this same combination of NFC parameters might be beneficial for the recognition of high-frequency consonants, such as the /s/ in panel (c). As shown by the spectrograms, the frication spanning 4.0–9.1 kHz in the original wideband signal [panel (a)] is shifted by NFC to the 2.5–3.3 kHz range [panel (c)]. While it appears as if this new signal has few discerning features useful for perception, the time waveform indicates that a significant amount of broadband temporal envelope information is preserved. Lowered temporal envelope information from frication can inform the hearing aid user about (a) the presence of a high-frequency speech event; (b) its subclass (a fricative) based on its noisy sound quality; (c) the relative duration of the original sound; and (d) the relative intensity and changes therein over the duration of event (i.e., the amplitude envelope).

Finally, when implementing any frequency-lowering algorithm for a hearing aid user, the upper frequency limit of aided audibility (i.e., the “maximum audible output frequency”) may be critical because it gives clinicians information about the frequency range that should be targeted for lowering (Alexander, 2014; Kimlinger *et al.*, 2015). In order to control the maximum output BW for the entire sample group, the experiments reported here used low-pass filtering

at 3.3 kHz or 5.0 kHz to restrict the aidable BW to limits representative of either a moderately severe (MS) high-frequency SNHL or a mild to moderate (MM) high-frequency SNHL. These two output BWs were examined because different outcomes were hypothesized based on the amount of information outside the audible BW and the subsequent frequency range available for recoding it. That is, while there is more speech information to be gained from frequency lowering as SNHL increases, the audible bandwidth available to recode the missing information also becomes increasingly narrower. In turn, frequency lowering becomes progressively more difficult to implement without distorting critical aspects of the low-frequency speech signal, such as formant frequencies. On the other hand, while individuals with less severe hearing loss have fewer speech perception deficits, they also have a wider bandwidth of audibility available to implement frequency lowering. This property combined with NFC processing may make it possible to provide some benefit to these individuals with less risk of distorting important low-frequency information (Alexander *et al.*, 2014; Brennan *et al.*, 2014; Picou *et al.*, 2015; Wolfe *et al.*, 2015).

II. METHOD

A. Listeners

Thirty native-English speaking adults with mild to moderate SNHL through 4 kHz completed the study. Tympanometry and bone conduction thresholds in each test ear were consistent with normal middle ear function (i.e., compliance peak height ≥ 0.3 mmho with peak pressure between -150 and $+50$ daPa, and most air-bone gaps ≤ 10 dB). Listeners were divided into two groups primarily on the basis of their high-frequency thresholds. All listeners in group MS had thresholds ≥ 60 dB hearing level (HL) at 6 and 8 kHz [Fig. 3(a)] and all but 3 listeners in group MM had thresholds ≤ 60 dB HL at these frequencies [Fig. 3(b)]. One listener from each group was excluded from data analysis due to unusually low scores in the wideband practice

FIG. 3. Audiometric thresholds for the listeners in group MS (a) and group MM (b), with each line and symbol combination representing a different listener and with the thick black line representing the group mean. The gray-shaded regions represent where stimuli were low-pass filtered at 3.3 kHz to create a moderately severe bandwidth restriction (a) or at 5.0 kHz to create a mild-to-moderate bandwidth restriction (b).

condition, especially on the high-frequency stimuli (described below) which were near chance performance ($\leq 15\%$). Group MS listeners (6 males, 8 females; age range = 47–83 years, median = 70 years) identified stimuli that were low-pass filtered at 3.3 kHz to create a moderately severe (MS) bandwidth restriction. All but one listener had thresholds ≤ 60 dB HL through 3 kHz, which was close to the maximum amplified frequency as indicated by the gray shading in Fig. 3. Three listeners from this group withdrew from the study before completing all of the conditions and are not reported in the total. All group MM listeners (7 males, 7 females; age range = 27–82 years, median = 62 years), who identified stimuli that were low-pass filtered at 5.0 kHz to create a mild-to-moderate (MM) bandwidth restriction,

had thresholds ≤ 60 dB HL through the maximum amplified frequency.

B. Speech material

1. Consonant stimuli

Consonant recognition was tested using 60 vowel-consonant-vowel (VCV) syllables formed by combining 20 consonants (/p, t, k, b, d, g, f, θ, s, ʃ, v, z, dʒ, tʃ, m, n, l, r, w, y/) and three vowels (/a, i, u/). The 60 VCVs were recorded with 16-bit resolution and 22.05 kHz sampling rate from three adult male and three adult female talkers, all with upper Midwestern accents. One male and one female talker were used for practice/familiarization (120 tokens) and the remaining talkers were used for testing (240 tokens). VCVs were mixed with speech-shaped noise at 10 dB SNR, which was generated by filtering white noise to match the 1/3-octave spectra of the international long-term average speech spectrum from Byrne *et al.* (1994), which represents the average of male and female talkers from 12 different languages.

2. Vowel stimuli

Vowel recognition was tested using the twelve /h/-vowel-/d/ (hVd) syllables from the Hillenbrand *et al.* (1995) database (/i, ɪ, e, ε, æ, a, ɔ, o, u, ʊ, ʌ, ɜ/). Eighteen talkers (6 men, 6 women, 3 boys, and 3 girls) from the 139-talker database (upper Midwestern accents) were selected. Based on data from 20 normal-hearing adults listening in quiet (Hillenbrand *et al.*, 1995), the chosen talkers all had overall identification rates of at least 97.5% correct and individual token identification rates of at least 90% correct. Six talkers (2 men, 2 women, 1 boy, and 1 girl) were used for practice/familiarization (72 tokens) and the remaining talkers were used for testing (144 tokens). The tokens were upsampled from 16.0 to 22.05 kHz and mixed with speech-shaped noise at 5 dB SNR. A lower SNR was used for vowel recognition than for consonant recognition because previous experience with these stimuli (Alexander *et al.*, 2011) indicated that a greater amount of noise would be needed to induce a sufficient number of errors so that perceptual differences between test conditions could be identified.

3. High-frequency stimuli

Stimuli for this experiment were chosen to maximize the relative information contained in the high-frequency speech spectrum. Fricative (/s, z, ʃ, f, v, θ, ð/) and affricate consonants (/tʃ, dʒ/) as spoken in a VC context by 3 female talkers 5 times each were used (cf. Stelmachowicz *et al.*, 2001; Alexander *et al.*, 2014). The vowel /i/ was used to minimize cues provided by formant transitions. Tokens were downsampled from 44.1 to 22.05 kHz and mixed with speech-shaped noise at 10 dB SNR. These cumulative steps resulted in stimuli where the primary spectral differences were confined to frequencies above 3 kHz (see Fig. 1 in Alexander *et al.*, 2014). One of the renditions for each VC and talker combination was used for practice/familiarization

(27 tokens) and the remaining four were used for testing (108 tokens).

C. Procedure

For each group, speech stimuli were processed in eight different conditions, described in detail in the signal processing section. The first condition tested was used for practice and always had a wideband frequency response, with individualized amplification through 8 kHz. The purpose was to expose listeners to the stimuli and to give them practice with the procedure before presenting them with altered stimuli; therefore, results for this condition are not considered further. The remaining seven conditions, referred to as “NFC conditions” included six conditions with NFC processing and one control condition without NFC that had been low-pass filtered to match the output BW of the other six conditions. The order of the conditions was counterbalanced across listeners using a partial Latin square design. For each condition, listeners first identified the consonant stimuli followed by the vowel stimuli. With breaks taken at the listeners’ discretion, each condition took about 45–60 min to complete. Listeners completed one or two conditions per experimental session. Following testing of the consonant and vowel stimuli, listeners were tested on the high-frequency stimuli in one or two sessions using the same order for the eight conditions. Each condition for this phase of testing took about 10–15 min to complete. Listeners completed half or all of the conditions for the high-frequency stimuli in one session.

Stimuli were presented in random order, monaurally with 16-bit resolution and 22.05 kHz sampling rate through BeyerDynamic DT150 headphones. Thirteen listeners used their left ear and 15 used their right ear for testing. Test ear selection was determined by the experimenter using the criteria described earlier. Every stimulus set had a practice stage with visual response feedback followed by a test stage without feedback. Listeners were instructed at the beginning of every session orally and once more at the beginning of each stimulus set with written instructions. A grid of all the orthographic response alternatives was displayed on a touchscreen computer monitor. Consonant stimuli were arranged alphabetically in a 4×5 grid using the letter of the medial consonant (e.g., “b”) for the orthographic representation. Vowel stimuli were arranged in a 4×3 grid and ordered according to tongue height (low to high) advancement and (front to back). The orthographic representation consisted of the “word” represented by the acoustic stimulus, and in order to avoid confusion by unfamiliar or non-words, it also consisted of a secondary word that shared the same vowel sound as the target: “had” (“hat”), “hod” (“hot”), “hawed” (“taught”), “head” (“bed”), “hayed” (“hate”), “heard” (“hurt”), “hid” (“hit”), “heed” (“heat”), “hoed” (“road”), “hood” (“hook”), “hud” (“hut”), and “who’d” (“hoot”). The high-frequency stimuli were arranged in a 3×3 grid in alphabetical order: “eeF,” “eeCH,” “eeJ,” “eeS,” “eeSH,” “eeH (teeth),” “eeTH (breathe),” “eeV,” “eeZ.” After a stimulus was presented, listeners indicated their response by using the touchscreen monitor or computer mouse to click on the place on the grid that corresponded to

their response. During the practice stage, if an incorrect response was given, the correct response briefly flashed red twice to cue the listener, otherwise, the response chosen briefly faded into the background and reappeared.

D. Signal processing

In order to have control and flexibility over the signal processing, a hearing aid simulator designed by the author in MATLAB (see also, Alexander and Masterson, 2015; McCreery *et al.*, 2013; McCreery *et al.*, 2014; Brennan *et al.*, 2014; Brennan *et al.*, 2016) was used to apply NFC processing to the speech and to amplify it using wide dynamic range compression (WDRC) and a LynxTWO™ sound card. For NFC-processed stimuli, signals were frequency compressed using an algorithm based on Simpson *et al.* (2005) before WDRC amplification. First, the input signal was divided into a low-pass band and a high-pass band using a 256-tap finite impulse response (FIR) filter and the SF as the crossover frequency. The low-pass band was not processed and was time aligned with the processed high-pass band after frequency lowering to form the final composite signal. The high-pass signal was processed using overlapping blocks of 256 samples. Blocks were windowed in the time domain using the product of a Hamming and a sinc function. Magnitude and phase information at 32-sample intervals (1.45 ms) were obtained by submitting this 256-point window to the “spectrogram” function in MATLAB with 224 points of overlap (7/8) and with 128 points for the discrete Fourier transforms. The instantaneous frequency, phase, and magnitude of each frequency bin above the SF, spanning a region up to a pre-chosen input BW, was used to modulate sine-wave carriers with frequency-reassignment determined from Eq. (1) in Simpson *et al.* (2005):

$$F_{out} = [SF^{1-1/CR}] \times [F_{in}^{(1/CR)}], \quad (1)$$

where F_{out} is the output frequency in kHz, SF is the start frequency kHz, CR is the frequency compression ratio, and F_{in} is the instantaneous input frequency in kHz. For each NFC condition, the SF and input BW were chosen to equal the frequency of one of the fast Fourier transform (FFT) bins, which were multiples of 172.26 Hz (22 050 sampling rate/128 FFT rate). For the same SF, higher input BWs necessarily came at the expense of higher CRs. The frequency input-output functions and the NFC parameters used to derive them for group MS and group MM are shown in panels (a) and (b) of Fig. 4, respectively. Different shades of gray are used to plot the input-output function for each SF, which is the same shading used for plotting the results. Because of the restricted output BW (3.3 kHz) for group MS, only two SFs (1.6 and 2.2 kHz) were tested, but three input BWs (5.0, 7.1, and 9.1 kHz) were able to be tested due to the wide frequency range of unamplified speech. Because group MM had a higher output BW (5.0 kHz), three SFs (1.6, 2.8, and 4.0 kHz) were tested, but only two input BWs (7.1 and 9.1 kHz) were tested due to the relatively narrower frequency range of unamplified speech. It should be noted that because the native sampling rate of the vowel stimuli was 16 kHz

FIG. 4. The frequency input-output functions and the NFC parameters used to derive them for group MS and group MM (top and bottom panels, respectively). Each SF is represented by a different shade of gray. For each SF, the desired input BWs were achieved by the selection of CR (see inset for values).

that the highest input BW was technically 8 kHz, not 9.1 kHz. Given, the essentially nil contribution of these high frequencies to vowel perception, this difference was not expected to influence the results.

Speech stimuli with and without NFC processing were processed with individualized amplification using methods outlined in Alexander and Masterson (2015), McCreery *et al.* (2013), McCreery *et al.* (2014), Brennan *et al.* (2014), and Brennan *et al.* (2016). First, absolute threshold in dB sound pressure level (SPL) at the tympanic membrane for each listener was estimated using a transfer function of the TDH 50 earphones (used for the audiometric testing) on a DB100 Zwislocki Coupler and KEMAR (Knowles Electronic Manikin for Acoustic Research). Threshold values and the number of compression channels (8) were entered in the Desired Sensation Level (DSL) m(I/O) v5.0 a algorithm for adults (National Centre for Audiology, University of Western Ontario). Individualized prescriptive values for compression threshold and compression ratio for each channel as well as target values for the real-ear aided responses were generated for a BTE hearing aid style with wide dynamic range compression, no venting, and a monaural fitting. Individualized prescriptions were generated using

8 nominal channels, each with a 5-ms attack time and a 50-ms release time. Gain for each channel was automatically tuned to targets using the “carrot passage” from Audioscan[®] (Dorchester, ON). Specifically, the 12.8 s speech signal was analyzed using 1/3-octave filters specified by ANSI (2004). A transfer function of the BeyerDynamic DT150 headphones on a DB100 Zwislocki Coupler and KEMAR was used to estimate the real-ear aided response. The resulting long term average speech spectrum was then compared with the prescribed DSL m(I/O) v5.0 a targets for a 60 dB SPL presentation level. The 1/3-octave values for dB SPL were grouped according to which WDRC channel they overlapped. The average difference for each group was computed and used to adjust the gain in the corresponding channel. Maximum gain was limited to 55 dB and minimum gain was limited to 0 dB (after accounting for the headphone transfer function). Because of filter overlap and subsequent summation in the low frequency channels, gain for channels centered below 500 Hz was decreased by an amount equal to $10\log_{10}(N \text{ channels})$, where $N \text{ channels}$ is the total number of channels used in the simulation.

Within MATLAB, stimuli were first scaled so that the input speech level was 60 dB SPL and then band-pass filtered into eight channels. Center and crossover frequencies were based on the recommendations of the DSL algorithm: 0.315, 0.5, 0.8, 1.25, 2.0, 3.15, and 5.0 kHz. Except for the wideband practice condition, the highest-frequency channel (5 kHz to Nyquist, 11.25 kHz) was assigned 0 dB of gain since the signal energy in this region was beyond the output BWs of the NFC conditions. The signal within each channel was smoothed using an envelope peak detector computed from Eq. (8.1) in Kates (2008). This set the attack and release times so that the values corresponded to what would be measured acoustically using ANSI (2003).

Output levels were set according to the level of the smoothed gain-control signal and a three-stage input/output function: unity gain up to the compression threshold, followed by WDRC, followed by compression limiting. Output compression limiting with a 10:1 compression ratio, 1-ms attack time, and 50-ms release time was implemented in each channel to keep the estimated real-ear aided response from exceeding the DSL-recommended broadband output limiting targets (BOLT) or 105 dB SPL, whichever was lower. The signals were summed across channels and subjected to a final stage of output compression limiting (10:1 compression ratio, 1-ms attack time, and 50-ms release time) to control the final presentation level and prevent saturation of the soundcard and/or headphones. Because the DSL algorithm for adults often prescribes expansion (compression ratio <1.0) in low frequency regions of normal hearing, this resulted in a 2-stage input/output function with negative gain up to the expansion threshold, followed by unity gain (compression limiting was functionally inactive due to the controlled presentation levels).

III. RESULTS

Statistical analyses were conducted with percent correct transformed to rationalized arcsine units (Studebaker, 1985)

in order to normalize the variance across the range. *Post hoc* testing of paired comparisons was completed using paired *t*-tests and the false discovery rate (FDR) correction of Benjamini and Hochberg (1995) for multiple comparisons. For statistically significant main effects and interactions, effect size is reported as the generalized eta-squared statistic, η_G^2 (Bakeman, 2005). For purposes of data reduction, consonants were grouped according to manner of articulation since this feature provides a broad classification of the acoustic composition of the different consonants (e.g., Kent and Read, 2002) that could influence recognition of VCVs processed with NFC (an analysis of individual VCV tokens is provided at the end of Sec. III). Similarly, in order to evaluate whether vowel acoustics, namely formant spacing, influenced recognition of hVds processed with NFC, vowels were analyzed by talker group and by phoneme so that patterns in the results could be examined in terms of relative formant frequency.

A. Consonant stimuli

1. Group MS: Moderately severe bandwidth restriction

A within-subjects analysis of variance (ANOVA) was performed with manner and NFC condition (here and throughout, this factor includes the low-pass filtered control or “NFC off” condition) as within-subjects factors. ANOVA results are shown Table I. The main effect for manner was significant with the following rank order: affricates [$M = 72.6\%$, $SE = 5.2\%$], nasals [$M = 64.2\%$, $SE = 2.5\%$], approximants [$M = 58.1\%$, $SE = 3.4\%$], stops [$M = 57.7\%$, $SE = 2.5\%$], and fricatives [$M = 49.5\%$, $SE = 3.4\%$]. All manners of articulation were significantly different from one another at the $p < 0.001$ level, except for the paired comparison between approximants vs stops which was significant at the $p < 0.05$ level.

The effects of NFC condition varied by manner as indicated by the significant interaction. To examine the interaction, separate ANOVAs and *post hoc* tests were performed for each manner (see Table II). There were no significant paired comparisons for affricates and only two for nasals, for which recognition in the 2.2 kHz SF with 5.0-kHz BW condition was significantly better than the 1.6-kHz SF conditions with 5.0 kHz and 7.1 kHz BWs. Paired comparisons for the remaining manners of articulation are illustrated in Fig. 5 using separate panels for the fricatives, stops, and approximants. Bars are grouped and shaded by the SF for NFC: the first set of light gray bars correspond to the three input BWs with a 1.6 kHz SF, the second set of medium-gray bars

TABLE I. Within-subjects omnibus ANOVA results for the effects of manner of articulation and processing condition (“Cond”) on VCV recognition in group MS. DF_n and DF_d denote the degrees of freedom in the numerator and denominator, respectively. Here and throughout, for each significant result, the generalized eta-squared statistic for effect size is denoted η_G^2 .

Effect	DF_n	DF_d	F	P	η_G^2
Cond	6	78	11.9	<0.001	0.1
Manner	4	52	27.3	<0.001	0.52
Manner $\times$ Cond	24	312	4.4	<0.001	0.11

TABLE II. Within-subjects ANOVA ($df = 6,78$) results for the effects of NFC condition on VCV recognition for each manner of articulation in group MS. Manners of articulation are arranged in order (high to low) according to the importance of high-frequency cues.

Vowel	F	p	η_G^2
Fricatives	12.8	<0.001	0.20
Stops	21.6	<0.001	0.23
Approximants	9.0	<0.001	0.18
Nasals	4.1	0.001	0.12
Affricates	1.8	0.10	

correspond to the three input BWs with a 2.2 kHz SF, and the last dark-gray bar corresponds to the 3.3-kHz low-pass control condition. Statistically significant paired comparisons ($p < 0.05$ after FDR correction) are indicated by the lines and symbols above the bars. For each line, the shaded dot indicates the reference condition used for comparison and, to the left it, subsequent shaded asterisks indicate conditions from which it is significantly different. The lines and symbols are shaded the same as the bars for each representative condition.

Significant paired comparisons for approximants were confined to the three NFC conditions with a 1.6 kHz SF, each of which had significantly lower recognition than the control condition and their respective 2.2-kHz SF condition with the equivalent input BW. The pattern of significant paired comparisons for stops was similar to that for approximants. However, recognition for the two 2.2-kHz SF conditions with the greatest amount of frequency compression (7.1 and 9.1 kHz BWs) were significantly lower than the control condition and recognition significantly decreased as the BW of frequency compression increased. For fricatives, significantly lower recognition occurred for the 1.6-kHz SF conditions with 7.1 and 9.1 kHz BWs. For the 2.2 kHz SF, recognition was significantly higher with 5.0 kHz BW than with 7.1 kHz and 9.1 kHz BWs, the latter of which was significantly lower than the control condition.

In summary, when analyzed by manner of articulation, most consonant categories, except affricates and nasals, showed an effect of NFC processing. For fricatives, stops, and approximants, the pattern of results was more similar than different with most of the significant detriments to recognition involving the 1.6 kHz SF. For this SF, the size of the detriment tended to increase as input BW increased, such that the 7.1- and 9.1-kHz BW conditions consistently had significantly lower recognition than the low-pass control condition. Viewed another way, for conditions with an equivalent input BW but a different SF, recognition was significantly better with a 2.2 kHz than a 1.6 kHz SF.

2. Group MM: Mild-to-moderate bandwidth restriction

Just as with group MS, a within-subjects ANOVA was performed on the data from group MM using manner and condition as within-subjects factors (see Table III). The main effect for manner was significant with recognition significantly better for affricates [$M = 77.0\%$, $SE = 3.9\%$] than approximants [$M = 68.5\%$, $SE = 2.9\%$], fricatives

FIG. 5. Mean recognition scores and *post hoc* test outcomes for consonants when analyzed by manner of articulation (different panels) for group MS. Bars are grouped and shaded by the SF (1.6 kHz and 2.2 kHz), with the labels underneath corresponding to the three input BWs (5.0, 7.1, and 9.1 kHz). The last dark-gray bar corresponds to the 3.3-kHz low-pass control condition. Statistically significant paired comparisons are indicated by the lines and symbols above the bars, with the shaded dot corresponding to the reference condition used for each comparison and the shaded asterisks corresponding the level of significance (* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$).

TABLE III. Within-subjects omnibus ANOVA results for the effects of manner of articulation and processing condition (“Cond”) on VCV recognition in group MM.

Effect	DF _n	DF _d	F	p	η_G^2
Cond	6	78	6.0	<0.001	0.06
Manner	4	52	6.0	<0.001	0.21
Manner $\times$ Cond	24	312	1.7	0.02	0.04

[$M = 67.6\%$, $SE = 2.9\%$], stops [$M = 67.0\%$, $SE = 2.6\%$], and nasals [$M = 64.1\%$, $SE = 3.2\%$], all at the $p < 0.001$ level. The only other significant paired comparison was between the approximants and the nasals [$p < 0.05$].

The effect of condition varied by manner as indicated by the significant interaction. To examine the interaction, separate ANOVAs and *post hoc* tests were performed for each manner (see Table IV). Figure 6 plots the mean recognition scores and results of the *post hoc* tests for the different manners of articulation in separate panels, except for affricates and nasals, which had no significant paired comparisons. All of the significant paired comparisons demonstrated lower recognition for one or both of the conditions with 1.6 kHz SF, especially the 9.1 kHz BW, which had slightly greater frequency compression than the 7.1 kHz BW (albeit, with the second smallest CR among the 12 conditions tested). Fricative recognition for the 1.6 kHz SF with 9.1 kHz BW was significantly lower than three of the other NFC conditions (both 4.0-kHz SF conditions and the 2.8 kHz SF with 7.1 kHz BW). Likewise, stop recognition for the 1.6 kHz SF with 9.1 kHz BW was significantly lower than all of the other conditions, including the 1.6 kHz SF with 7.1 kHz BW, which was only significantly lower than the 4.0 kHz SF with 9.1 kHz BW. Recognition for approximants was significantly lower for both 1.6 kHz-SF conditions compared to the low-pass control condition, the 4.0 kHz SF with 7.1 kHz BW condition, and to one of the 2.8-kHz SF conditions each.

In summary, analyses of VCV recognition for group MM were similar to group MS because the fricatives, stops, and approximants, showed a pattern in which all of the significant detriments to recognition involved the 1.6 kHz SF, even though the amount of frequency compression (CR) was less for group MM. For this SF, detriment was more likely with the 7.1 kHz BW than with the 9.1 kHz BW.

TABLE IV. Within-subjects ANOVA ($df = 6,78$) results for the effects of NFC condition on VCV recognition for each manner of articulation in group MM. Manners of articulation are arranged in order (high to low) according to the importance of high-frequency cues.

Vowel	F	p	η_G^2
Fricatives	4.8	0.003 ^a	0.08
Stops	6.0	<0.001	0.08
Approximants	6.0	<0.001	0.11
Nasals	2.0	0.07	
Affricates	1.1	0.39	

^a p -values after correction using the Greenhouse-Geisser epsilon for within-subjects factors with $p < 0.05$ on Mauchly’s test for sphericity.

FIG. 6. Mean recognition scores and *post hoc* test outcomes for consonants when analyzed by manner of articulation (different panels) for group MM. Bars are grouped and shaded by the SF (1.6 kHz, 2.8 kHz, and 4.0 kHz), with the labels underneath corresponding to the two input BWs (7.1 and 9.1 kHz). The last dark-gray bar corresponds to the 5.0-kHz low-pass control condition.

B. Vowel stimuli

1. Group MS: Moderately severe bandwidth restriction

A within-subjects ANOVA was performed with talker group, vowel, and NFC condition as within-subjects factors.

TABLE V. Within-subjects omnibus ANOVA results for the effects of talker group (“Talker”), vowel, and processing condition (“Cond”) on hVd recognition in group MS.

Effect	DF _n	DF _d	F	p	η_G^2
Talker	2	26	18.4	<0.001	0.02
Vowel	11	143	24.0	<0.001	0.36
Cond	6	78	15.8	<0.001	0.05
Talker × Vowel	22	286	7.6	<0.001	0.06
Talker × Cond	12	156	2.5	0.004	0.01
Vowel × Cond	66	858	3.8	<0.001	0.06
Talker × Vowel × Cond	132	1716	1.2	0.10	

As shown in Table V, all main effects were significant and were qualified by significant two-way interactions, indicating that the effects of NFC depended on the talker group and on the vowel. The three-way interaction between these factors was not significant.

To examine the interaction between talker group and condition, separate ANOVAs and *post hoc* tests were performed for each talker group (see Table VI). The mean recognition scores for each talker group are plotted in different panels in Fig. 7. Across talker groups, all of the significant differences between conditions involved the 1.6 kHz SF. Furthermore, there was a systematic pattern across talker groups in terms of the involvement of the different input BWs with this SF. For children, all three BWs were significantly lower than the low-pass control condition and were significantly lower than each of the 2.2-kHz SF conditions, with one exception. For women, only the 7.1 and 9.1 kHz BWs showed significant detriment compared to the other conditions. For men, most of the significant comparisons were confined to the 9.1 kHz BW condition.

To examine the interaction between vowel and condition, separate ANOVAs and *post hoc* tests were performed for each vowel. As indicated in Table VII, seven of the twelve vowels showed a significant ANOVA result. Vowels that did not show a significant effect of NFC were characterized by a low F_2 (/o, u, u/) or a high F_1 (/æ, a/). Furthermore, *post hoc* tests for each significant ANOVA revealed that /ɔ/ (low F_2) and /I/ (high F_2) had no significant pairs after controlling for multiple comparisons. As noted by Hillenbrand *et al.* (1995), /ɔ/ stands out from the rest of the vowel set and is a high source of confusion because many speakers (and listeners) of American English do not distinguish between /ɔ/ and /a/. This was also true in the present study because recognition for the unprocessed, low-pass control condition for /ɔ/ was 47.6% ($SE = 7.3\%$) compared to

TABLE VI. Within-subjects ANOVA ($df = 6,78$) results for the effects of NFC condition on hVd recognition for each talker group in group MS.

Vowel	F	p	η_G^2
Children	16.5	<0.001 ^a	0.31
Women	15.6	<0.001 ^a	0.30
Men	5.8	<0.001	0.12

^ap-values after correction using the Greenhouse-Geisser epsilon for within-subjects factors with $p < 0.05$ on Mauchly’s test for sphericity.

FIG. 7. Mean recognition scores and *post hoc* test outcomes for vowels when analyzed by talker for group MS. Results are plotted in the same manner as Fig. 5 for children, women, and men in the top, middle, and bottom panels, respectively.

56.5% ($SE=6.8\%$) for /a/, 77.3% ($SE=5.3\%$) for /æ/, and $>88\%$ for the nine remaining vowels ($M=95.0\%$, $SE=1.3\%$).

To evaluate whether the pattern of differences between NFC conditions varied among the five vowels with significant paired comparisons (/ɜ̃, ʌ, ɛ, e, i/), a within-subjects

ANOVA was performed on the reduced data set. The interaction between vowel and NFC condition was significant [$F(24, 312)=2.2$, $p=0.001$, $\eta_G^2=0.06$]. Further inspection revealed that /i/ was responsible for the significant interaction because an ANOVA with the remaining vowels only revealed no significant interaction [$F(18, 234)=1.6$, $p=0.060$]. Based on these analyses, the mean recognition scores and *post hoc* results for the vowels /ɜ̃, ʌ, ɛ, e/ are plotted in the top panel of Fig. 8 and the results for /i/ are plotted in the bottom panel. Just as with the consonants, almost all of the significant paired comparisons involved the 1.6-kHz SF conditions, especially the 7.1 and 9.1 kHz BWs, which had significantly lower recognition than each of the other conditions (including the 1.6 kHz SF with 5.0 kHz BW). Whereas, recognition for the 1.6 kHz SF with 5.0 kHz BW (the leftmost bar) was significantly lower than the low-pass control and two of the conditions with the 2.2 kHz SF for vowels in the top panel, it was not significantly lower for /i/ (bottom panel). An examination of the two plots suggests that this difference was likely due to differences in the variances and not the means associated with each of the comparisons. Another difference between the patterns of results for these two vowel sets is the size of the detriment associated with NFC. Compared to the low-pass control, recognition for the 1.6 kHz SF with 7.1 and 9.1 kHz BWs was about 15% to 20% lower for /ɜ̃, ʌ, ɛ, e/ and about 25% to 35% lower for /i/.

Figure 9 illustrates the effects of NFC on the acoustic vowel space for the different conditions. Using the formant frequencies at “steady state” as reported by Hillenbrand *et al.* (1995), each data point represents the F_1 , F_2 frequency coordinates for the twelve vowels, averaged across all 12 talkers. The black circles represent the low-pass control condition, while the light-gray and dark-gray symbols represent the 1.6- and 2.2-kHz SF conditions, respectively. Triangles, squares, and diamonds represent the 5.0, 7.1, and 9.1 kHz BW conditions, respectively, and the horizontal dotted lines correspond to the two SFs (vowels with average F_2 frequencies below the SFs show minor shifts because some individual tokens had F_2 frequencies that exceeded the SFs). As

TABLE VII. Within-subjects ANOVA ($df = 6,78$) results for the effects of NFC condition on hVd recognition for each vowel in group MS. Vowels are arranged in phonetic order according to tongue position from front to back (generally, high to low second formant frequency).

Vowel	F	p	η_G^2
i	20.2	<0.001	0.37
ɪ	3.5	0.004	0.11
e	16.8	<0.001	0.40
ɛ	8.9	<0.001	0.32
æ	0.8	0.60	
ʌ	6.4	<0.001	0.18
ɜ̃	10.4	<0.001	0.32
a	1.6	0.17	
u	0.3	0.92	
ʊ	0.7	0.65	
o	1.6	0.17	
ɔ	2.7	0.02	0.05

FIG. 8. For group MS, mean recognition scores and *post hoc* test outcomes for the vowels /ɜ, ʌ, ɛ, e/ are shown in the top panel and recognition scores and *post hoc* test outcomes for the vowel /i/ are shown in the bottom panel. Analyses revealed that only these vowels demonstrated significant differences between conditions and that the effect of condition was statistically equivalent for the vowels represented in the top panel.

shown in the figure, vowels with the highest F_2 frequencies in the low-pass condition have the greatest shifts in F_2 following NFC. Changes in SF cause greater shifts in F_2 compared to changes in BW (compare the locations of the light-gray and dark-gray symbols), which seems to explain the general pattern of consonant and vowel recognition across NFC conditions. The Pearson correlation across all vowel tokens indicated that the size of the detriment in vowel recognition associated with NFC (re: LPC) was significantly related to the change in F_2 frequency (re: LPC) [$r(850) = -0.38, p < 0.001$].

In summary, vowels with F_2 frequencies above the SF were most affected by NFC. Therefore, as with consonants, the only significant differences involved detriments to recognition for the 1.6-kHz SF conditions. Compared to the 2.2-kHz SF, changes in input BW with this SF resulted in greater shifts in F_2 frequency. The higher the F_2 frequency before lowering (e.g., children vs men talkers, vowel /i/ vs vowels

FIG. 9. Acoustic vowel space showing the effects of nonlinear frequency compression on the formant frequencies for the twelve vowels, averaged across all talkers in group MS. The light-gray and dark-gray symbols represent the 1.6- and 2.2-kHz SF conditions, respectively, and the black circles represent the low-pass control condition. Triangles, squares, and diamonds represent the 5.0-, 7.1-, and 9.1-kHz BW conditions, respectively. The two horizontal dotted lines represent the two start frequencies. The smaller, open square and diamond symbols, respectively, represent the 1.6 kHz SF with 7.1- and 9.1-kHz BW conditions from group MM because these were the only conditions from this group that caused a noticeable change in average F_2 frequency. Note that the 1.6 kHz SF with 5.0 kHz BW (CR = 1.57) from group MS (light-gray triangles) and the 1.6 kHz SF with 9.1 kHz BW (CR = 1.52) from group MM (small, open diamonds) had almost the same frequency remapping functions, with the exception that the latter extended to 5.0 kHz in the output.

/ɜ, ʌ, ɛ, e/), the greater the shift and the greater the effect of input BW on vowel recognition.

2. Group MM: Mild-to-moderate bandwidth restriction

For the vowel stimuli, fewer significant differences were expected compared to group MS because of the higher SFs used for group MM. Just as for group MS, a within-subjects ANOVA was performed with talker group, vowel, and NFC condition as within-subjects factors. As shown in Table VIII, all main effects were significant and were qualified by at least one significant two-way interaction. While the two-way interaction between vowel and condition was significant, the two-way interaction between talker group and condition was not significant. The three-way interaction between all of these factors was not significant.

To examine the interaction between vowel and condition, separate ANOVAs and *post hoc* tests were performed for each vowel. As indicated in Table IX, only four (/ɜ, ʌ, ɛ, e/) of the twelve vowels showed a significant effect of NFC condition. These four vowels also showed significant effects for the listeners in group MS. A within-subjects ANOVA on

TABLE VIII. Within-subjects omnibus ANOVA results for the effects of talker group (“Talker”), vowel, and processing condition (“Cond”) on hVd recognition in group MM.

Effect	DF _n	DF _d	F	p	η_G^2
Talker	2	26	21.1	<0.001	0.01
Vowel	11	143	22.5	<0.001	0.37
Cond	6	78	3.2	0.007	0.01
Talker × Vowel	22	286	4.0	<0.001	0.04
Talker × Cond	12	156	1.6	0.08	
Vowel × Cond	66	858	1.4	0.04	0.02
Talker × Vowel × Cond	132	1716	0.9	0.79	

the reduced data set comprised of only these four vowels indicated that there was no interaction between NFC condition and vowel [$F(18, 234) = 1.2, p = 0.26$]. The mean recognition scores and *post hoc* outcomes for the vowels /ɜ, ʌ, ε, e/ are plotted in Fig. 10 and show that recognition for the 1.6 kHz SF with 9.1 kHz BW was significantly lower than all other conditions and that no other paired comparisons were significantly different.

Acoustic analyses confirmed that the effect of NFC on the shift in F_2 frequency was almost exclusively confined to the 1.6 kHz SF since the 2.8 kHz SF was out of the range of all but the highest F_2 frequencies and the 4.0 kHz SF was completely out of the F_2 range. The mean shifts in F_2 frequency across talkers for the 1.6 kHz-SF conditions are represented in Fig. 9 as the smaller, open symbols. Shifts in F_2 frequency associated with the 9.1 kHz BW (open diamonds) were essentially the same as the 1.6 kHz SF with 5.0 kHz BW condition for group MS (light-gray triangles) because both conditions shared the same SF and roughly the same CR (1.52 and 1.57, respectively), with the only difference being the range of maximum input and output frequencies. Shifts in F_2 frequency associated with the 7.1 kHz BW (open squares) tended to overlap with the 2.2 kHz SF conditions for group MS (dark-gray symbols) or fall between the conditions represented by the two SFs. Because of the relatively small shifts in F_2 frequency associated with the 1.6 and 2.8 kHz SFs, the Pearson correlation between all F_2 shifts with these two SFs and the change in vowel

TABLE IX. Within-subjects ANOVA (df = 6,78) results for the effects of NFC condition on hVd recognition for each vowel in group MM.

Vowel	F	p	η_G^2
i	1.7	0.13	
ɪ	0.7	0.66	
e	4.6	<0.001	0.21
ɛ	4.4	<0.001	0.15
æ	1.7	0.13	
ʌ	2.8	0.017	0.06
ɜ	3.7	0.003	0.14
a	0.7	0.68	
u	0.4	0.89	
ʊ	0.9	0.51	
o	0.9	0.53	
ɔ	0.9	0.49

FIG. 10. For group MM, mean recognition scores and *post hoc* test outcomes for the vowels /ɜ, ʌ, ε, e/. Analyses revealed that only these vowels demonstrated significant differences between conditions and that the effect of condition was statistically equivalent across these vowels.

recognition (re: LPC) was only marginally significant [$r(566) = -0.08, p = 0.06$].

In summary, only a limited number of vowels were negatively affected by the NFC settings used for group MM. The vowels that were affected were the same subset of vowels that showed significant differences between conditions in group MS, excluding /i/. Furthermore, consistent with the results from group MS, the NFC setting that resulted in the greatest shifts in F_2 frequency, the 1.6-kHz SF with 9.1 kHz BW, was also associated with the greatest detriments to recognition.

C. High-frequency stimuli

1. Group MS: Moderately severe bandwidth restriction

For the fricatives and affricates spoken in the /iC/ context (henceforth denoted “HF phonemes” for “high-frequency phonemes”), a within-subjects ANOVA was performed with HF phoneme and NFC condition as within-subjects factors (see Table X). The main effect for HF phoneme was significant. The highest recognition occurred with /ʃ/ [$M = 90.3\%, SE = 4.2\%$], and with the affricates, /tʃ/ [$M = 89.2\%, SE = 4.0\%$] and /dʒ/ [$M = 80.3\%, SE = 7.8\%$], which were each significantly better than the remaining fricatives at the $p < 0.001$ level. Recognition for /ʃ/ and /tʃ/ was not significantly different from each other ($p = 0.31$), but each was significantly better than /dʒ/ ($p < 0.001$ and $p = 0.005$, respectively). The lowest recognition occurred with /ð/ [$M = 20.1\%, SE = 4.8\%$], which was significantly

TABLE X. Within-subjects ANOVA results for the effects of phoneme and processing condition (“Cond”) on recognition of the high-frequency stimuli in group MS.

Effect	DF _n	DF _d	F	p	η_G^2
Cond	6	78	2.9	0.01	0.02
Phoneme	8	104	27.3	<0.001	0.50
Phoneme × Cond	48	624	3.3	<0.001	0.10

worse than each of the other HF phonemes at the $p < 0.001$ level. The rank order of recognition for the remaining HF phonemes was /v/ [$M = 64.9\%$, $SE = 3.3\%$], /z/ [$M = 56.7\%$, $SE = 5.2\%$], /s/ [$M = 50.1\%$, $SE = 6.7\%$], /θ/ [$M = 51.4\%$, $SE = 6.3\%$], and /f/ [$M = 44.0\%$, $SE = 5.8\%$], with the only significant comparisons ($p \leq 0.01$) occurring for pairs between nonadjacent ranks, except /z/-/θ/ and /s/-/f/.

The main effect for condition was also significant. As shown in the top panel of Fig. 11, unlike the data shown for the VCVs and hVds, almost all of the NFC conditions had recognition scores that were significantly higher than the low-pass control. The effect of NFC condition also depended on the HF phoneme as indicated by the significant interaction. To examine the interaction, separate within-subjects ANOVAs and *post hoc* analyses were performed for each HF phoneme, which revealed that /s/ [$F(6,78) = 6.7$, $p < 0.001$, $\eta_G^2 = 0.20$] and /z/ [$F(6,78) = 5.8$, $p < 0.001$, $\eta_G^2 = 0.21$] were the only HF phonemes with significant differences between NFC conditions. The middle and bottom panels in Fig. 11 show the mean recognition scores and results of the *post hoc* tests for /s/ and /z/, respectively. The pattern of results is radically different from those for VCVs and hVds because recognition for the low-pass control condition was almost 50% poorer for /s/ and almost 45% poorer for /z/ relative to the best conditions with NFC processing, which were often the worst conditions for the VCVs and hVds. /s/ recognition for the low-pass control condition was significantly lower than each of the conditions with NFC, except the 2.2 kHz SF with 7.1 and 9.1 kHz BWs. The highest /s/ recognition occurred with the 1.6 kHz SF with 7.1 and 9.1 kHz BWs, which were not significantly different from each other, but were significantly higher than each of the other conditions, with two exceptions for the 9.1 kHz BW. For /z/ recognition, the only significant paired comparisons were between the low-pass control condition and the 1.6 kHz SF with 7.1 and 9.1 kHz BWs and the 2.2 kHz SF with 7.1 kHz BW. The latter condition was also significantly higher than the 2.2 kHz SF with 5.0 kHz BW, suggesting that listeners benefited from the additional input BW, despite the higher compression ratio necessary to bring it into the frequency range of the output BW.

2. Group MM: Mild-to-moderate bandwidth restriction

A within-subjects ANOVA was performed with HF phoneme and NFC condition as within-subjects factors (see Table XI). The main effect for HF phoneme was significant. The rank order of recognition was /f/ [$M = 90.1\%$, $SE = 6.7\%$], /tʃ/ [$M = 89.7\%$, $SE = 3.6\%$], /dʒ/ [$M = 83.0\%$, $SE = 8.0\%$], /s/ [$M = 82.7\%$, $SE = 6.1\%$], /z/ [$M = 82.7\%$, $SE = 3.2\%$], /v/ [$M = 69.8\%$, $SE = 6.3\%$], /θ/ [$M = 59.1\%$, $SE = 3.4\%$], /f/ [$M = 58.5\%$, $SE = 5.7\%$], and /ð/ [$M = 29.0\%$, $SE = 4.3\%$]. All paired comparisons were statistically significant ($p < 0.05$), except /ʃ/-/tʃ/, /dʒ/-/s/-/z/, and /θ/-/f/.

The main effect for condition was also significant. As shown in the top panel of Fig. 12, three of the four conditions with 2.8 or 4.0 kHz SF had recognition scores that were significantly higher than the low-pass control condition,

FIG. 11. Mean recognition scores and *post hoc* test outcomes for the high-frequency stimuli in group MS. The overall mean is plotted in the top panel and the individual results for /s/ and /z/ are plotted in the middle and bottom panels, respectively.

while the two conditions with 1.6 kHz SF did not show a difference. The effect of NFC condition depended on the HF phoneme as indicated by the significant interaction. To examine the interaction, separate within-subjects ANOVAs and *post hoc* analyses were performed for each HF phoneme, which revealed that /f/ [$F(6,78) = 4.1$, $p = 0.001$,

TABLE XI. Within-subjects ANOVA results for the effects of phoneme and processing condition (“Cond”) on recognition of the high-frequency stimuli in group MM.

Effect	DF _n	DF _d	F	p	η_G^2
Cond	6	78	8.7	<0.001	0.03
Phoneme	8	104	26.6	<0.001	0.52
Phoneme × Cond	48	624	3.9	<0.001	0.11

$\eta_G^2=0.09$], /s/ [$F(6,78)=10.3$, $p<0.001$, $\eta_G^2=0.19$], and /z/ [$F(6,78)=13.9$, $p<0.001$, $\eta_G^2=0.36$] were the only HF phonemes with significant differences between NFC conditions. *Post hoc* analyses for /f/ recognition indicated that the only significant paired comparison involved the NFC condition with the lowest recognition, the 1.6 kHz SF with 7.1 kHz BW [$M=42.9\%$, $SE=5.7\%$], and the NFC condition with the highest recognition, the 2.8 kHz SF with 9.1 kHz BW [$M=72.6\%$, $SE=8.1\%$]. The middle and bottom panels in Fig. 12 show the mean recognition scores and results of the *post hoc* tests for /s/ and /z/, respectively. Relative to the control condition, /s/ and /z/ recognition was significantly improved by NFC processing, just as with group MS. However, unlike group MS, significant benefit was observed for each of the six conditions with NFC processing, which were not significantly different from each other. Across conditions, NFC improved /s/ and /z/ recognition by approximately 33%–43%.

D. Microanalysis of NFC effects on VCV recognition using individual tokens

Useful information about the perceptual effects of NFC processing may be obscured by the need to average across speech features/classes in order to have a manageable presentation of the data. That is, each listener was tested on almost 500 speech tokens per condition, almost half of which were VCVs that were analyzed by averaging the data within manner class and across gender and vowel context. Toscano and Allen (2014) make arguments for the need to account for acoustic and perceptual variability at the level of individual speech tokens in order to develop a better diagnostic understanding of speech perception deficits by listeners with SNHL. They emphasize that two productions of the same phoneme spoken by two different talkers [or spoken in two different phonetic contexts] may vary in their perceptual robustness to noise, for example. Furthermore, they suggest that most error distributions tend to be bimodal; such that some phonemes have little to no errors while other phonemes have large errors associated with them. This line of reasoning naturally extends to frequency-lowering hearing aids, which can alter the natural relationship between speech cues originating from different parts of the spectrum.

In order to evaluate which of the 240 VCV tokens contributed most to the differences observed across the NFC conditions, they were ranked ordered according to the amount of variance each individual token contributed to the variance of the entire stimulus set. This was done separately for the data from the two groups of listeners. In order to have a manageable set for the sake of discussion, only tokens that

FIG. 12. Mean recognition scores and *post hoc* test outcomes for the high-frequency stimuli in group MM. The overall mean is plotted in the top panel and the individual results for /is/ and /iz/ are plotted in the middle and bottom panels, respectively.

contributed 25% cumulative variance from group MS (13 tokens) and group MM (15 tokens) are presented in Tables XII and XIII, respectively. These relatively small sets represent about 5% of the total number of VCV tokens. In fact, around 50% of the total variance for each group could be explained by only 40 of the 240 tokens. Interestingly, for

TABLE XII. For group MS, the rank ordering of the 13 VCVs that contributed the greatest variance (cumulative total 25%) to the overall dataset is displayed. The two male (m1 and m2) and two female (f1 and f2) talkers are listed in the “Talker” column. All talkers had upper Midwestern dialects with the following approximate fundamental frequencies: f1 = 190 Hz, f2 = 220 Hz, m1 = 120 Hz, m2 = 90 Hz. The consonant and vowel making up the VCV are listed in the “C” and “V” columns, respectively. The percent of listeners (n = 14) who correctly identified each token is listed in the columns corresponding to the 1.6 start frequency (SF), the 2.2 kHz SF, and the control condition, with the header beneath each corresponding to the input BW (5.0, 7.1, and 9.1 kHz). The last column displays the cumulative variance accounted for by each successive entry in the table.

Talker	C	V	1.6 kHz SF			2.2 kHz SF			Control 3.3 kHz	Cumulative Variance
			5.0 kHz	7.1 kHz	9.1 kHz	5.0 kHz	7.1 kHz	9.1 kHz		
m1	/k/	a	64.3%	0.0%	42.9%	100.0%	92.9%	85.7%	92.9%	3.3%
f2	/r/	a	64.3%	28.6%	7.1%	78.6%	85.7%	85.7%	100.0%	6.1%
f2	/θ/	a	92.9%	21.4%	7.1%	64.3%	28.6%	21.4%	42.9%	8.3%
m2	/s/	u	35.7%	7.1%	0.0%	78.6%	50.0%	42.9%	71.4%	10.5%
m1	/t/	u	100.0%	64.3%	21.4%	92.9%	92.9%	85.7%	100.0%	12.5%
f2	/z/	i	50.0%	28.6%	28.6%	57.1%	14.3%	14.3%	92.9%	14.5%
m1	/s/	u	57.1%	14.3%	14.3%	85.7%	50.0%	50.0%	78.6%	16.4%
m1	/ʃ/	a	78.6%	35.7%	28.6%	85.7%	50.0%	35.7%	92.9%	18.2%
f1	/dʒ/	i	42.9%	71.4%	92.9%	28.6%	64.3%	71.4%	21.4%	19.9%
f2	/k/	a	35.7%	50.0%	42.9%	71.4%	85.7%	92.9%	100.0%	21.5%
f2	/k/	i	7.1%	14.3%	7.1%	50.0%	7.1%	42.9%	64.3%	23.0%
m2	/t/	i	50.0%	14.3%	7.1%	57.1%	57.1%	64.3%	64.3%	24.4%
m1	/z/	i	85.7%	85.7%	42.9%	92.9%	92.9%	92.9%	42.9%	25.8%

group MS, 10 of the top 24 tokens contained /s/ or /z/ (5 each), with all 4 of the /usu/ tokens being represented. For group MM, 8 of the top 15 tokens contained /s/, /z/, or /ʃ/, with 3 of the 4 /aʃa/ tokens being represented.

The data presented in Tables XII and XIII are revealing in at least three ways. First, the variability in recognition across conditions for individual tokens indicates that a small change in SF or BW can make a substantial difference in recognition. For example, the second entry in Table XII (f2/ara/) shows that 100% listeners identified this token correctly in the control condition, whereas only 1 of them (7.1%) identified it correctly with the 1.6 kHz SF and 9.1 kHz BW. However, keeping SF the same while decreasing BW to 5.0 kHz increased recognition to 64.3% and keeping BW the same

while increasing SF increased recognition to 85.7%. The following examples further support this observation as well as a second observation from the data: *no one condition simultaneously maximizes recognition for all the tokens*. Increasing input BW from 5.0 to 9.1 kHz (by increasing the CR) with the 1.6 kHz SF for group MS (Table XII) resulted in a drop in the recognition rate from 93% to 7% for f2/aθa/ and from 100% to 21% for m1/utu/. However, this same change in settings resulted in an *increase* in the recognition rate from 43% to 93% for f1/idʒi/. For group MM (Table XIII), increasing input BW from 7.1 to 9.1 kHz with the 1.6 kHz SF resulted in a drop in the recognition rate from 71% to 14% for f1/ibi/, but resulted in an increase in the recognition rate from 14% to 64% f1/apa/. The third observation is that there are cases with

TABLE XIII. Similar to Table XII, displayed for group MM (n = 14) is the rank ordering of the 15 VCVs that contributed the greatest variance (cumulative total 25%) to the overall dataset (240 tokens) is displayed.

Talker	C	V	1.6 kHz SF		2.8 kHz SF		4.0 kHz SF		Control 5.0 kHz	Cumulative Variance
			7.1 kHz	9.1 kHz	7.1 kHz	9.1 kHz	7.1 kHz	9.1 kHz		
f1	/k/	i	21.4%	7.1%	71.4%	50.0%	85.7%	64.3%	71.4%	3.0%
f2	/z/	u	21.4%	21.4%	64.3%	78.6%	57.1%	64.3%	85.7%	5.3%
f1	/s/	a	42.9%	7.1%	57.1%	50.0%	78.6%	64.3%	78.6%	7.4%
m2	/ʃ/	a	100.0%	100.0%	100.0%	92.9%	71.4%	92.9%	35.7%	9.5%
f1	/f/	i	21.4%	0.0%	50.0%	35.7%	50.0%	42.9%	71.4%	11.3%
f1	/b/	i	71.4%	14.3%	50.0%	57.1%	71.4%	78.6%	71.4%	13.1%
f2	/ʃ/	a	42.9%	0.0%	42.9%	35.7%	57.1%	64.3%	21.4%	14.7%
f1	/ʃ/	a	85.7%	85.7%	92.9%	71.4%	71.4%	100.0%	35.7%	16.3%
f1	/t/	u	57.1%	64.3%	78.6%	35.7%	85.7%	78.6%	92.9%	17.7%
m1	/t/	i	85.7%	85.7%	85.7%	57.1%	85.7%	42.9%	50.0%	19.0%
m2	/s/	u	78.6%	28.6%	78.6%	71.4%	85.7%	78.6%	78.6%	20.4%
f1	/ʃ/	i	78.6%	71.4%	85.7%	92.9%	78.6%	85.7%	35.7%	21.6%
m2	/w/	i	42.9%	35.7%	78.6%	78.6%	71.4%	42.9%	50.0%	22.8%
f2	/p/	i	71.4%	100.0%	57.1%	64.3%	92.9%	92.9%	57.1%	24.0%
f1	/z/	u	85.7%	50.0%	78.6%	100.0%	100.0%	92.9%	100.0%	25.2%

the same syllable (e.g., /izi/ in Table XII) where NFC impairs performance for one talker (f2) but improves it for another (m1).

IV. DISCUSSION

This study examined how different ways of compressing inaudible mid- and/or high-frequency information at lower frequencies influences the perception of consonants and vowels embedded in isolated syllables. For this reason, low-pass filtering and the selection of NFC parameters was used to fix the output bandwidth (BW) at 3.3 or 5.0 kHz, respectively, representing a moderately severe (MS) or a mild-to-moderate (MM) BW restriction, respectively. The effects of different combinations of NFC start frequency (SF) and input BW (by varying the compression ratio, CR) were examined using two groups of listeners, one for each output BW.

For both groups of listeners, the most consistent finding across the analyses for consonants in a VCV context and vowels in an hVd context was a significant decrease in recognition for one or more of the conditions with the lowest SF, 1.6 kHz, relative to the low-pass control conditions. With a few exceptions, speech recognition for conditions with higher SFs for group MS (2.2 kHz) and group MM (2.8 and 4.0 kHz) were not significantly different from the low-pass control conditions. It was also common for one or more conditions with the 1.6 kHz SF to have significantly lower recognition compared with another condition with an equivalent input BW but with a higher SF (consequently, a higher CR), especially for group MS. In addition, among the 1.6-kHz SF conditions the likelihood of a significant detriment to speech recognition increased as the CR (input BW) increased. With a few exceptions (e.g., stops and fricatives in VCVs for group MS), this was not true at the higher SFs, where fewer significant differences were observed across input BW.

The effects of manipulating CR to create the desired input BWs are consistent with the importance of preserving low-frequency information over maintaining spectral resolution of the lowered speech. Recall that for a fixed output BW and a fixed SF that input BW only could be increased by increasing the CR. Stating the relationship another way, within a listener group (fixed output BW) in order to compress an equivalent BW of information from the input using a higher SF, CR necessarily had to be increased. Therefore, findings of significantly better vowel and consonant recognition for conditions with a SF higher than 1.6 kHz but with an equivalent input BW (higher CR) suggests that the primary concern for audiologists when manipulating CR should not be how much frequency compression there is, but instead the concern should be *where* the frequency compression occurs. This is supported by the finding that there were no significant differences for group MM between the 2.8 kHz and 4.0 kHz SF when comparing across input BW and vice versa.

The pattern of results described in the preceding paragraph, which is also consistent with some of the studies discussed in the Introduction, suggest that caution should be warranted when using low SFs, especially with higher CRs. Using normal-hearing listeners and a 1.6 kHz SF, Ellis and

Munro (2013) found that compared to a control condition without NFC, recognition for sentences in noise was significantly worse with NFC. Souza *et al.* (2013) found minimal effects of NFC on sentence recognition when SF = 2.0 kHz, regardless of CR, and when CR = 1.5, regardless of SF, including 1.5 kHz. Adverse effects of NFC were most evident for the 1.0 kHz SF with CR = 2.0 and 3.0. The finding that the 1.5 kHz SF with 1.5 CR had minimal effects on sentence in noise recognition is at odds with the negative findings for most of the 1.6 kHz-SF conditions in this study, in which 3 of the 5 conditions across both groups had CRs ≤ 1.57 . As noted in the Introduction, there were several differences between the Souza *et al.* (2013) study and the other studies mentioned, including this one, in terms of their methods used for frequency compression (e.g., the frequency remapping function for each CR was unknown) and for amplification (linear). In addition, differences in the test material between Souza *et al.* (sentences with varying amounts of noise) and this study (syllables with fixed noise) also could have contributed to the observed differences with the low SF. Finally, in terms of perceived speech and music quality, Parsa *et al.* (2013) found that listeners were influenced more by manipulations of SF than of CR, with sound quality ratings gradually decreasing as SF decreased below 3.0 kHz, especially for normal-hearing listeners in quiet.

From the previous discussion, it is clear that low SFs have an increased potential for negatively influencing speech recognition, however, the frequency or frequency range where this begins is unclear. Based on the data from this study, it appears that the negative effects begin somewhere between 1.6 and 2.2 kHz. It is well documented in the speech perception literature that frequencies of formant peaks and transitions play a critical role in distinguishing co-articulated phonemes because they provide important information about the shape of the vocal tract during speech production (e.g., Ladefoged and Broadbent, 1957; Pols *et al.*, 1969; Klein *et al.*, 1970; Sussman *et al.*, 1991; Dorman and Loizou, 1996; Bachorowski and Owren, 1999; Hillenbrand and Nearey, 1999; Neel, 2008). Therefore, it is speculated that negative effects of NFC were most evident for these SFs because they represented the range where many phonemes contained all or a portion of their second formant energy. For example, the finding that changes in input BW had the most effect on recognition with the 1.6 kHz SF, with fewer significant differences between input BWs at higher SFs suggests that an increase in frequency compression (CR) was most deleterious when F_2 was in the region of lowering.

The hVd results for talker group and vowel were also consistent with the hypothesis that F_2 frequency plays an integral role in the effect of NFC on speech recognition. Specifically, for group MS recognition of vowels spoken by child talkers was affected more by NFC than vowels spoken by women talkers, with recognition of vowels spoken by men talkers being the least affected. This pattern of results fits with speech acoustics since F_2 frequency co-varies with vocal tract length, being higher for children (short vocal tracts) and lower for men (longer vocal tracts). Furthermore, individual vowels with the lowest F_2 frequencies (cf. Fig. 9) were unaffected by NFC. However, it was not true that all

vowels with mid or high F_2 frequencies were affected by NFC, despite undergoing relatively large shifts changes in frequency (/I, æ/ for both groups, and /i/ for group MM). One possible explanation for this contradictory finding is that the proximity of other phonemes in the vowel space (i.e., how similar the frequency-lowered vowel is to its neighbors) and the availability of other cues, such as duration which helps distinguish between tense and lax vowels sharing similar places (e.g., /e/ vs /I/), mediate the importance of F_2 on recognition frequency-lowered vowels (Pols *et al.*, 1969; Klein *et al.*, 1970; Neel, 2008).

An analysis of VCV recognition in terms of manner of articulation also provides broad support for the role of F_2 frequency in the perception of NFC-processed speech. Consonants were analyzed by manner of articulation because this property tends to segregate consonants into classes with distinct acoustic properties (e.g., Kent and Read, 2002). For example, recognition of affricates and nasals was least affected by NFC for both group MS and group MM, arguably because these consonants depended the least on F_2 . That is, affricates can often be distinguished from their constituent phonemes, stops and fricatives, using temporal cues like amplitude rise time and duration and nasals can often be distinguished from other phonemes by their low first formant frequency and the presence of a low-frequency anti-formant. There was no clear difference in the results as a function of NFC for consonants with the remaining manners of articulation (fricatives, stops, and approximants), likely because discrimination of phonemes within each manner were equally dependent on lower-frequency information provided by the formant transitions (versus higher-frequency information in the frication for the fricatives and stops).

It might be surprising that few analyses demonstrated clear improvements in recognition with NFC. At outset of the study, no improvement in vowel recognition was expected because of the frequency specificity of their perceptual cues as discussed in the preceding paragraphs. Instead, the hVd stimuli were included to assess whether improvements in consonant recognition would come at a cost in the form decrements in vowel recognition. This expectation was partially supported by the data because 33%–50% improvements in consonant recognition for the /s/ and /z/ from the high-frequency stimulus set and for isolated tokens from the VCV stimulus set (mainly, fricative and stop consonants) occurred with NFC settings that were the most detrimental to low-frequency dominated vowels (low SF, medium to high CR). These results are consistent with several other studies on children and adults that found that improvements with NFC were primarily limited to medial or final /s/ and /z/ in quiet or in noise, especially using tests of discrimination threshold and plural detection (e.g., Simpson *et al.*, 2005; Wolfe *et al.*, 2010; Wolfe *et al.*, 2011; Wolfe *et al.*, 2015; Glista *et al.*, 2009; Glista *et al.*, 2012; Alexander *et al.*, 2014; Hopkins *et al.*, 2014; Zhang *et al.*, 2014; Picou *et al.*, 2015; Ellis and Munro, 2015). The difference in outcomes for the final vs medial /s/ in this study (VCV context vs /iC/ context) may be attributed to non-spectral cues that can help distinguish it from the final /ʃ/, such as duration and intensity (e.g., Jongman *et al.*, 2000),

with the latter being longer and louder. In addition, because formant transitions precede and follow the medial /s/ and because they are lower in frequency for /ʃ/ than for /s/, VCV conditions with significant frequency compression in second and third formant range may have created additional acoustic cues that more closely match the phonetic profile of /ʃ/ than of /s/, thereby leading to an increased number of confusions.

At least two limitations of the current study may have limited the improvement in consonant recognition with NFC: linguistic context and experience. There is no doubt that listeners would benefit from the additional linguistic information provided by sentences, however, it is unclear how this would affect relative benefit from NFC. On the one hand, additional context might help listeners overcome the low-level perceptual deficits by allowing them to fill in the missing high-frequency information, thereby leaving little room for improvement from frequency lowering. On the other hand, contextual cues might aid the use/learning of NFC-processed speech and consequently minimize some of the observed side effects, such as the aforementioned /s/-/ʃ/ confusions. Only a few studies have examined the relative performance of NFC using both consonant and sentence recognition tests. The available data suggests that the effect of context might depend on the severity of loss and consequently on the need for additional information beyond that provided by contextual cues. Testing listeners with mild to moderate SNHL, both Wolfe *et al.* (2015) and Picou *et al.* (2015) found improvements in /s/ discrimination by children and adults, respectively, following ~1-month trial periods with NFC-enabled hearing aids, but found no statistical improvements on tests of speech recognition in noise. On the other hand, Wolfe *et al.* (2011) reported improvements on both consonant recognition and sentence recognition in noise for children with moderate SNHL after 6 months of regular use of hearing aids with NFC, whereas early testing at 6 weeks did not show a significant difference for the sentence recognition task (Wolfe *et al.*, 2010). Ellis and Munro (2015) also reported statistical improvements on tests of consonant recognition and sentence recognition in noise for adults with moderately severe SNHL, and in contrast to the Wolfe *et al.* (2011) study, they found immediate improvement in recognition with further improvement over 6 weeks limited to consonant recognition.

As just highlighted by the previous discussion (Wolfe *et al.*, 2011; Ellis and Munro, 2015), there is very little evidence to suggest who and under what circumstances benefit from NFC will improve after extended experience with the processing. Whether listeners can perceptually recalibrate to altered formant patterns over time with is beyond the present discussion, but it seems clear from these experiments that acute exposure without training to frequency-shifted formants can be detrimental to speech recognition. With respect to the design of the current study, it is perhaps more likely than studies that used actual hearing aids, that the lack of experience with the different NFC processing strategies was detrimental to the perception of low-level speech cues and consequently suppressed recognition of consonants processed with NFC compared to the control conditions. This is possible because listeners only had practice trials

with which to become acquainted to the new speech codes introduced by the different NFC settings, which were switched after about an hour when testing consonant and vowel recognition and after about 15 min when testing the high-frequency stimuli. That is, listeners never had an opportunity to perceptually adjust to the new settings by listening to more complex linguistic structures or, importantly, to the sound of their own voice. This may have been a greater hindrance to those conditions with the most frequency lowering, namely, the 1.6 kHz-SF conditions, compared to conditions with milder NFC settings. Given these limitations, the results of this study might be best construed as a “worse-case scenario” in which NFC with SFs ≥ 2.2 kHz did little harm to the perception of most phonemes.

V. CONCLUSIONS

- (1) Outcomes when assessing benefit from NFC depend in large part on its setting(s) and the sounds used to evaluate it. Improvements in vowel recognition should not be expected. At minimum, improvements in final /s/ and /z/ identification can be expected when using NFC for a variety of hearing losses.
- (2) Use of a start frequency (SF) less than 2.2 kHz may degrade speech understanding, especially when combined with a high compression ratio (CR). In cases where the bandwidth of audibility is restricted, it is better to increase SF at the expense of a higher CR, than vice versa. In cases where the bandwidth of audibility is less restricted, attempts should be made to keep the SF above the frequency range of most second formants.
- (3) While increasing the CR reduces the spectral resolution of the lowered signal, the most important factor is where the compression occurs and not how much of it there is. When the SF is low, increasing CR decreases identification for many consonants and vowels. A slightly lower CR can be maintained by bringing less high-frequency information down into the range of audibility (lower input BW). On the other hand, if the reduction in input BW is too great, final /s/ and /z/ identification might not be optimized (see #4). For higher SFs, if sufficient input BW is made audible, CR seems to be less important.
- (4) No one set of parameters simultaneously maximizes recognition for all tokens; the “optimal” settings for individual phonemes may vary according to their acoustic characteristics. That is, settings that maximize the identification of one phoneme might come at the expense of a decrease identification of one or more phonemes. Because a hearing aid is not just worn for select tokens, if NFC is to benefit speech recognition, it must improve the overall mean, which represents the many different types of tokens the listener will encounter. Therefore, in practice, one may need to compromise on optimal performance for fricatives to avoid a decrease in recognition for the vowels and other consonants. It is possible, but untested, whether ideal performance for these settings will be reached after a period of learning/acclimatization. Alternatively, perhaps a future implementation of NFC can balance these tradeoffs by applying frequency

lowering periodically based on the features that it is able to enhance, while leaving other time intervals of speech unaltered. In this vein, the information presented in this paper will be useful for refining the precision of the signal processing to selectively target those stimuli for which it can provide benefit.

ACKNOWLEDGMENTS

The nonlinear frequency compression component of the hearing aid simulator would not have been possible without the programming efforts of Minseok Kwon and program testing by the Hearing and Amplification Research Lab at Boys Town National Research Hospital. Three anonymous reviewers provided very useful suggestions on an earlier version of this manuscript. This research was funded by NIDCD Grant No. RC1 DC010601.

- Alexander, J. M. (2010). “Maximizing benefit from nonlinear frequency compression,” in *4th Phonak Virtual Audiology Conference*.
- Alexander, J. M. (2013). “Individual variability in recognition of frequency-lowered speech,” *Semin. Hear.* **34**, 86–109.
- Alexander, J. M. (2014). “How to use probe microphone measures with frequency-lowering hearing aids,” *Audiol. Prac.* **6**(4), 8–13.
- Alexander, J. M., Jenison, R. L., and Kluender, K. R. (2011). “Real-time contrast enhancement to improve speech recognition,” *PLoS One* **6**(9), e24630.
- Alexander, J. M., Kopun, J. G., and Stelmachowicz, P. G. (2014). “Effects of frequency compression and frequency transposition on fricative and affricate perception in listeners with normal hearing and mild to moderate hearing loss,” *Ear Hear* **35**, 519–532.
- Alexander, J. M., and Masterson, K. (2015). “Effects of WDRC release time and number of channels on output SNR and speech recognition,” *Ear Hear* **36**, e35–e49.
- ANSI (2003). ANSI S3.22-2003, *Specification of Hearing Aid Characteristics* (American National Standards Institute, New York).
- ANSI (2004). ANSI S1.11-2004, *Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters* (American National Standards Institute, New York).
- Bachorowski, J., and Owren, M. (1999). “Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech,” *J. Acoust. Soc. Am.* **106**, 1054–1063.
- Baer, T., Moore, B. C. J., and Kluk, K. (2002). “Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies,” *J. Acoust. Soc. Am.* **112**, 1133–1144.
- Bakeman, R. (2005). “Recommended effect size statistics for repeated measures designs,” *Behav. Res. Methods* **37**, 379–384.
- Benjamini, Y., and Hochberg, Y. (1995). “Controlling the false discovery rate: A practical and powerful approach to multiple testing,” *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300.
- Brennan, M. A., McCreery, R., Kopun, J., Alexander, J. M., Lewis, D., and Stelmachowicz, P. G. (2014). “Paired comparisons of nonlinear frequency compression, extended bandwidth, and restricted bandwidth hearing-aid processing for children and adults with hearing loss,” *J. Am. Acad. Audiol.* **25**, 983–998.
- Brennan, M. A., McCreery, R., Kopun, J., Hoover, B., Alexander, J. M., Lewis, D., and Stelmachowicz, P. G. (2016). “Masking release in children with hearing loss when using amplification,” *J. Am. Acad. Audiol.* (in press).
- Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Nasser Kotby, M., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westerman, S., and Ludvigsen, C. (1994). “An international comparison of long-term average speech spectra,” *J. Acoust. Soc. Am.* **96**, 2108–2120.
- Dorman, M. F., and Loizou, P. C. (1996). “Relative spectral change and formant transitions as cues to labial and alveolar place of articulation,” *J. Acoust. Soc. Am.* **100**, 3825–3830.

- Ellis, R., and Munro, K. (2013). "Does cognitive function predict frequency compressed speech recognition in listeners with normal hearing and normal cognition?" *Int. J. Audiol.* **52**, 14–22.
- Ellis, R., and Munro, K. (2015). "Benefit from, and acclimatization to, frequency compression hearing aids in experienced adult hearing-aid users," *Int. J. Audiol.* **54**, 37–47.
- Glista, D., Scollie, S., Bagatto, M., Seewald, R., Parsa, V., and Johnson, A. (2009). "Evaluation of nonlinear frequency compression: Clinical outcomes," *Int. J. Audiol.* **48**, 632–644.
- Glista, D., Scollie, S., and Sulkers, J. (2012). "Perceptual acclimatization post nonlinear frequency compression hearing aid fitting in older children," *J. Speech Lang. Hear. Res.* **55**, 1765–1787.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Hillenbrand, J., and Nearey, T. (1999). "Identification of resynthesized /hVd/ utterances: Effects of formant contour," *J. Acoust. Soc. Am.* **105**, 3509–3523.
- Hopkins, K., Khanom, M., Dickinson, A., and Munro, K. (2014). "Benefit from non-linear frequency compression hearing aids in a clinical setting: The effects of duration of experience and severity of high-frequency hearing loss," *Int. J. Audiol.* **53**, 219–228.
- Jongman, A., Wayland, R., and Wong, S. (2000). "Acoustic characteristics of English fricatives," *J. Acoust. Soc. Am.* **108**, 1252–1263.
- Kates, J. M. (2008). *Digital Hearing Aids* (Plural, San Diego, CA), p. 233.
- Kent, R. D., and Read, C. (2002). *Acoustic Analysis of Speech*, 2nd ed. (Delmar Cengage Learning, Clifton Park, NY).
- Kimlinger, C., McCreery, R., and Lewis, D. (2015). "High-frequency audibility: The effects of audiometric configuration, stimulus type, and device," *J. Am. Acad. Audiol.* **26**, 128–137.
- Klein, W., Plomp, R., and Pols, L. C. W. (1970). "Vowel spectra, vowel spaces, and vowel identification," *J. Acoust. Soc. Am.* **48**, 999–1009.
- Ladefoged, P., and Broadbent, D. E. (1957). "Information conveyed by vowels," *J. Acoust. Soc. Am.* **29**, 98–104.
- McCreery, R., Venediktov, R., Coleman, J., and Leech, H. (2012). "An evidence-based systematic review of frequency lowering in hearing aids for school-age children with hearing loss," *Am. J. Audiol.* **21**, 313–328.
- McCreery, R. W., Alexander, J. M., Brennan, M. A., Hoover, B., Kopun, J., and Stelmachowicz P. G. (2014). "The influence of audibility and short-term audio-visual exposure on speech recognition with nonlinear frequency compression for children and adults with hearing loss," *Ear Hear* **35**, 440–447.
- McCreery, R. W., Brennan, M. A., Hoover, B., Kopun, J., and Stelmachowicz, P. G. (2013). "Maximizing audibility and speech recognition with nonlinear frequency compression by estimating audible bandwidth," *Ear Hear* **34**, e24–e27.
- Moore, B. C. J., Stone, M. A., Füllgrabe, C., Glasberg, B. R., and Puria, S. (2008). "Spectro-temporal characteristics of speech at high frequencies, and the potential for restoration of audibility to people with mild-to-moderate hearing loss," *Ear Hear* **29**, 907–922.
- Neel, A. (2008). "Vowel space characteristics and vowel identification accuracy," *J. Speech Lang. Hear. Res.* **51**, 574–585.
- Parsa, V., Scollie, S., Glista, D., and Seelisch, A. (2013). "Nonlinear frequency compression: Effects on sound quality ratings of speech and music," *Trends Amplif.* **17**, 54–68.
- Picou, E., Marcum, S., and Ricketts, T. (2015). "Evaluation of the effects of nonlinear frequency compression on speech recognition and sound quality for adults with mild to moderate hearing loss," *Int. J. Audiol.* **54**, 162–169.
- Pols, L., van der Kamp, L. J. Th., and Plomp, R. (1969). "Perceptual and physical space of vowel sounds," *J. Acoust. Soc. Am.* **46**, 458–467.
- Simpson, A. (2009). "Frequency-lowering devices for managing high-frequency hearing loss: A review," *Trends Amplif.* **13**(2), 87–106.
- Simpson, A., Hersbach, A. A., and McDermott, H. J. (2005). "Improvements in speech perception with an experimental nonlinear frequency compression hearing device," *Int. J. Audiol.* **44**, 281–292.
- Souza, P. E., Arehart, K., Kates, J., Croghan, N. E., and Gehani, N. (2013). "Exploring the limits of frequency lowering," *J. Speech Lang. Hear. Res.* **56**, 1349–1363.
- Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., and Lewis, D. E. (2001). "Effect of stimulus bandwidth on the perception of /s/ in normal and hearing-impaired children and adults," *J. Acoust. Soc. Am.* **110**, 2183–2190.
- Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., and Lewis, D. E. (2002). "Aided perception of /s/ and /z/ by hearing-impaired children," *Ear Hear.* **23**, 316–324.
- Stelmachowicz, P., Pittman, A., Hoover, B., Lewis, D., and Moeller, M. (2004). "The importance of high-frequency audibility in the speech and language development of children with hearing loss," *Arch. Otolaryngol. Head Neck Surg.* **130**, 556–562.
- Studebaker, G. A. (1985). "A rationalized arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Sussman, H., McCaffrey, H. A., and Matthews, S. A. (1991). "An investigation of locus equations as a source of relational invariance for stop place categorization," *J. Acoust. Soc. Am.* **90**, 1309–1325.
- Toscano, J., and Allen, J. (2014). "Across- and within-consonant errors for isolated syllables in noise," *J. Speech Lang. Hear. Res.* **57**, 2293–2307.
- Vickers, D. A., Moore B. C. J., and Baer T. (2001). "Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies," *J. Acoust. Soc. Am.* **110**, 1164–1175.
- Wolfe, J., Andrew, J., Schafer, E., and Nyffeler, N. (2011). "Long-term effects of non-linear frequency compression for children with moderate hearing loss," *Int. J. Aud.* **50**, 396–404.
- Wolfe, J., John, A., Schafer, E., Hudson, M., Boretzki, M., Scollie, S., Woods, W., Wheeler, S., Hudgens, K., and Neumann, S. (2015). "Evaluation of wideband frequency responses and non-linear frequency compression for children with mild to moderate high-frequency hearing loss," *Int. J. Audiol.* **54**, 170–181.
- Wolfe, J., John, A., Schafer, E., Nyffeler, M., Boretzki, M., and Caraway, T. (2010). "Evaluation of nonlinear frequency compression for school-age children with moderate to moderately severe hearing loss," *J. Am. Acad. Audiol.* **21**, 618–628.
- Zhang, V., Ching, T., Van Buynder, P., Hou, S., Flynn, C., Burns, L., McGhie, K., and Wong, A. (2014). "Aided cortical response, speech intelligibility, consonant perception and functional performance of young children using conventional amplification or nonlinear frequency compression," *Int. J. Ped. Otorhinolaryn.* **78**, 1692–1700.