Pitch and spectral resolution: A systematic comparison of bottom-up cues for top-down repair of degraded speech[a)]

Jeanne Clarke,[b)] Deniz Başkent,[c)] and Etienne Gaudrain[d)]
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, P.O. Box 30.001, BB21, 9700 RB Groningen, The Netherlands

(Received 3 April 2015; revised 21 October 2015; accepted 4 January 2016; published online 21 January 2016)

The brain is capable of restoring missing parts of speech, a top-down repair mechanism that enhances speech understanding in noisy environments. This enhancement can be quantified using the phonemic restoration paradigm, i.e., the improvement in intelligibility when silent interruptions of interrupted speech are filled with noise. Benefit from top-down repair of speech differs between cochlear implant (CI) users and normal-hearing (NH) listeners. This difference could be due to poorer spectral resolution and/or weaker pitch cues inherent to CI transmitted speech. In CIs, those two degradations cannot be teased apart because spectral degradation leads to weaker pitch representation. A vocoding method was developed to evaluate independently the roles of pitch and spectral resolution for restoration in NH individuals. Sentences were resynthesized with different spectral resolutions and with either retaining the original pitch cues or discarding them all. The addition of pitch significantly improved restoration only at six-bands spectral resolution. However, overall intelligibility of interrupted speech was improved both with the addition of pitch and with the increase in spectral resolution. This improvement may be due to better discrimination of speech segments from the filler noise, better grouping of speech segments together, and/or better bottom-up cues available in the speech segments. © *2016 Acoustical Society of America.*

[http://dx.doi.org/10.1121/1.4939962]

[EB] Pages: 395–405

I. INTRODUCTION

In everyday life, speech is often degraded by surrounding masking sounds and background noise before reaching the listener's ears. Yet normal-hearing (NH) listeners are, most of the time, still able to understand the message in such adverse listening situations. For that, NH listeners must restore speech segments that have been masked by the competing sound. *Phonemic restoration* (PR) is the ability of the brain to repair missing segments of speech (Warren, 1970) with use of linguistic knowledge, context and expectations (Samuel, 1981; Verschuure and Brocaar, 1983; Bashford *et al*., 1992). The *phonemic restoration effect* is measured by the increase in intelligibility of interrupted sentences when periodic silent interruptions are filled with noise bursts (e.g., Bashford *et al*., 1992; Başkent, 2012). The silent gaps introduced by interruptions may be misinterpreted as lexical cues, e.g., sudden starts or stops that can be interpreted as word boundaries, thereby affecting segmentation and speech rhythm. For example, if a silent interruption on the word "**cat**egory" would leave only "**cat**" heard, the listener would be more likely to wrongfully report the word "**cat**" because it was activated in their lexicon. Addition of noise bursts in the silent gaps may mask these spurious cues (Warren and Obusek, 1971), provided the noise is a potential masker (Bashford *et al*., 1992). As a result, filling the gaps with noise helps group the speech segments into a more continuous percept via segregation mechanisms (discrimination between speech and noise and sequential grouping of speech segments). Perhaps as a consequence, or in parallel, filling the gaps with noise also facilitates the lexical activation of the right words. When the silent interruptions are masked by noise, spurious word boundaries are not perceived anymore. The addition of the noise introduces an ambiguity that increases the number of possible words that can fit the degraded signal. For example, if noise would surround "**cat**" from "**cat**egory," the listener would activate "**cat**" and all the words embedding "**cat**," such as "**cat**egory" but also "**cat**erpillar," "**cat**fish," "s**cat**ter," "con**cat**enate," "un**cat**egorized," "meer**kat**," etc. Listeners are then given a broader choice of activated words in their lexicon; this increases the possibility that the correct lexical candidate is activated and restoration is facilitated (Srinivasan and Wang, 2005).

The aforementioned two aspects of sequential segregation, i.e., grouping and discrimination, are worth considering

[a)] Portions of this work were presented in "Is weak pitch fully responsible for low comprehension of interrupted speech with poor spectral resolution?," in Conference on Implantable Auditory Prostheses (CIAP'13), Lake Tahoe, CA, July 2013, and in "Top-down repair of speech: adding pitch to spectrally degraded speech," in Association for Research in Otolaryngology Midwinter Meeting (ARO'14), San Diego, CA, February 2014.

[b)] Also at: Research School of Behavioral and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, Groningen, The Netherlands. Electronic mail: j.n.clarke.kno@gmail.com

[c)] Also at: Research School of Behavioral and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, Groningen, The Netherlands.

[d)] Also at: Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics Team, CNRS UMR 5292, INSERM U1028, University Lyon 1, Lyon, France.

in phonemic restoration. When interruptions are left silent, the speech is perceived as a single stream that includes the silent gaps, which are then perceived as spurious cues that hinder intelligibility. When interruptions are filled with noise, the speech and the noise can be perceived as two distinct streams, and provided the noise is a plausible masker, the spurious cues from the silent gaps are masked, thus mitigating the loss of intelligibility. If stream segregation did not occur in the noise condition, the noise would also be integrated with the speech, also resulting in spurious cues of a different kind, potentially impairing intelligibility like in the silent condition. In the noise condition, sequential segregation relies on the spectral similarity of the successive speech segments (Singh, 1987) but also on the fundamental frequency (for tones, Moore and Gockel, 2012) in contrast to the noise segments, which have different spectral envelopes and periodicity properties. However, there seems to be a trade-off for the perceptual similarity between the speech segments and the noise bursts. For the noise to act like the most efficient masker, it has to be perceptually similar to speech (Bashford *et al.*, 1992), whereas stream segregation relies on the fact that speech and noise are perceptually different.

Cochlear implant (CI) users, for whom auditory bottom-up cues are degraded, as well as NH individuals listening to CI simulated speech, show positive phonemic restoration effects in fewer conditions than NH listeners (Başkent, 2012; Bhargava *et al.*, 2014). Başkent (2012) showed that the restoration benefit was only present at high spectral resolution conditions of the CI simulations. Bhargava *et al.* (2014) showed that for actual CI users, the restoration benefit was only present at longer duty-cycle speech conditions. Hence changes in acoustic bottom-up cues seem to induce changes in top-down repair of the speech, and this could be one of the factors contributing to speech perception difficulties CI users encounter in background noise (Fu and Nogaki, 2005; Stickney *et al.*, 2004). Illustrating this, some studies have shown that the addition of low-frequency speech information to the spectrally degraded (vocoded) speech, to simulate electro-acoustic stimulation (EAS), improved intelligibility of speech interrupted with silence and noise (Başkent, 2012; Başkent and Chatterjee, 2010). These improvements depended on the spectral resolution of the vocoded part.

The main forms of degradation of the bottom-up cues that occur due to the signal processing and signal transmission in CIs are reduced spectral resolution and weak pitch percept (for a review, see Rubinstein, 2004). In CIs, spectral resolution is limited by the number of electrodes and the extent to which current spreads around the stimulating electrode. Reducing the number of electrodes degrades the resolution of the spectral envelope as well as the spectral fine structure (i.e., the harmonic structure, which carries some pitch information). Pitch is the percept of a talker's fundamental frequency (F0) related to the periodicity of the signal. While in NH listeners pitch is encoded both through temporal and spectral mechanisms (Carlyon and Shackleton, 1994), in CIs, only temporal pitch cues are relatively preserved while spectral–or place–pitch cues are severely degraded (Moore and Carlyon, 2005). Despite the presence of periodicity information through temporal cues, the degradation of the harmonic structure strongly reduces pitch saliency. In CI devices (as well as in CI simulations), these two degradations are thus not independent. As both reduced resolution of the spectral envelope and degraded pitch happen together, we cannot tease apart which of these two factors causes the abnormal restoration performance observed in CI users.

Indeed the combination of both degradations could interfere with the discrimination between the noise interruption and the speech, making the noise as much a source of spurious cues as the silence (Bhargava *et al.*, 2014), which may explain, at least partly, the decrease of restoration when NH people listen to CI simulations. Reducing spectral resolution, by degrading the spectral envelope and making formant information less precise, may degrade phonetic cues to a degree that they become too ambiguous for restoration to happen. But pitch itself could also play a direct role in restoration. First, F0 cues could improve the intelligibility of the remaining speech segments as supported by studies that showed that whisper (without F0 cues) is less intelligible than voiced speech (for Japanese word recognition: Irino *et al.*, 2012; for consonants: Tartter, 1989, for vowels: Tartter, 1991; for concurrent syllable recognition: Vestergaard and Patterson, 2009). The expected benefit of F0 on intelligibility is also supported by the fact that F0 cues give information on voicing to distinguish voiced from unvoiced consonants (although other cues than F0 are also acoustically and perceptually correlated with voicing, such as loudness or duration, as shown in Peng *et al.*, 2012; Winn *et al.*, 2012). Second, the addition of low pass filtered speech to degraded speech (to simulate EAS) has been shown to improve restoration (Başkent, 2012). In line with other studies where the EAS benefit has been strongly associated with more available pitch cues (e.g., Brown and Bacon, 2009), Başkent (2012) argued that providing pitch information in the low-frequency part could help bind speech segments through temporal interruptions (as previously suggested by, for instance Plack and White, 2000). However, it was shown more recently that the EAS benefit can also be observed with addition of first formant (F1) information (Verschuur *et al.*, 2013), suggesting that the EAS benefit observed by Başkent (2012) for top-down restoration could not be only dependent on pitch. For Dutch vowels, F1 ranges from 259 Hz (for /u/) to 717 Hz (for /a/) (Adank *et al.*, 2004). The benefit observed in EAS simulation that adds speech low-pass filtered at 500 Hz may also come from some F1 cues along with the F0 information. Moreover, in a recent study, Clarke *et al.* (2014) showed that manipulating the average value of the F0 on speech segments across interruptions had no effect on speech intelligibility nor on phonemic restoration, indicating that pitch continuity is not necessary for restoration. The roles of pitch and spectral resolution for top-down restoration thus remain unclear.

In the present study, we have developed a new vocoding technique (based on TANDEM-STRAIGHT, Kawahara and Morise, 2011) to systematically and orthogonally vary the resolution of the spectral envelope (further referred to as "spectral resolution") and pitch availability (absence/presence) and to investigate which of these bottom-up cues accounted for the poorer top-down repair in CI simulated speech as well as with actual CIs. We hypothesized that adding F0 information to degraded speech would increase the phonemic restoration except at full spectral resolution where sufficient spectral detail is available to discriminate speech from noise. Moreover, adding F0 cues would also provide clearer speech features in the remaining degraded speech segments, such as strengthening lexical stress and sentence stress, enabling voice onset time use and voiced/unvoiced distinctions, at the linguistic level, and speaker normalization at the indexical level. Thus adding F0 cues could also increase the intelligibility of interrupted speech.

II. METHODS

A. Participants

Nineteen normal-hearing listeners, aged 19–36 yr (mean = 23.3, s.d. = 4.6), participated in the study. All participants were native speakers of Dutch, reporting no history of hearing or speech-related problems. Their pure-tone thresholds were 20 dB hearing level (HL) or less at audiometric frequencies between 250 and 6000 Hz for both ears. Written informed consent was collected from the listeners prior to their participation. Financial compensation was provided to the participants for their time. The study protocol was approved by the Medical Ethical Review Committee (Medisch Etische Toetsingscommissie) of the University Medical Center Groningen.

B. Stimuli

The lists of a corpus of 507 Dutch sentences spoken by a male talker were used (Versfeld *et al.*, 2000). Each list had 13 sentences, which were grammatically and syntactically correct and contained between four and nine words. The words were no longer than three syllables. The corpus was digitized at a 44.1 kHz sampling rate. The same sentence (Nos. 164 and 270), although uttered differently, appeared in two lists. Therefore list 13, containing sentence 164, was discarded.

C. Signal processing: TANDEM-STRAIGHT based vocoding

As an acoustic simulation of CIs, we used a new vocoding technique based on TANDEM-STRAIGHT (Kawahara and Morise, 2011), implemented in MATLAB. Specifically, this resynthesis technique allowed us to manipulate the voice in two independent ways: absence/presence of the original F0 (−F0 and +F0, respectively) and the resolution of the spectral envelope (simulating the number of bands in the CI simulation). The independent manipulation of these two parameters would not have been possible with a traditional CI simulation as the spectral degradation leads to weaker pitch representation. This new CI simulation provided a number of differences compared to traditional CI simulations. First, TANDEM-STRAIGHT does not implement channel interactions. Indeed we did not apply filters on each band (as commonly done in noise-band vocoders, such as by Shannon *et al.*, 1995), but instead we averaged the extracted spectral envelope per band (similar to applying a rectangular filter). Second, there are still some temporal F0 cues in common vocoders, whereas they are all removed with TANDEM-STRAIGHT. This latter point was confirmed by inspecting the processed stimuli (Fig. 2) as well as the auditory excitation patterns,[1] which are the output of the Auditory Image Model (AIM; Patterson, 2000). The AIM provides a representation of the sensory response to a sound through the peripheral auditory system (for more details, see Appendix A).

Figure 1 shows the steps of the offline processing of the voice manipulated in TANDEM-STRAIGHT (Kawahara and Morise, 2011), implemented in MATLAB. The speech signal was first decomposed into two parts: the spectral finestructure containing the F0 contour (the "source" element) and the spectral envelope (the "filter" element). An estimate for each of these elements was obtained every millisecond.

The spectral resolution was full, or reduced to 16, 8, 6, and 4 bands by manipulating the extracted spectral envelope. First, for all conditions, the spectral range was limited to 150–7000 Hz to match the range of most CIs and CI simulations (e.g., Başkent and Chatterjee, 2010). Then the spectral envelope was averaged per bands of frequencies. The band boundaries were chosen to have equal distances along the basilar membrane (Greenwood, 1990).

Independently of the manipulation on the spectral envelope, the F0 was kept unchanged for the +F0 conditions (with the original voiced and unvoiced parts) or set to zero for the −F0 conditions (which causes TANDEM-STRAIGHT to use a noise excitation as source). The

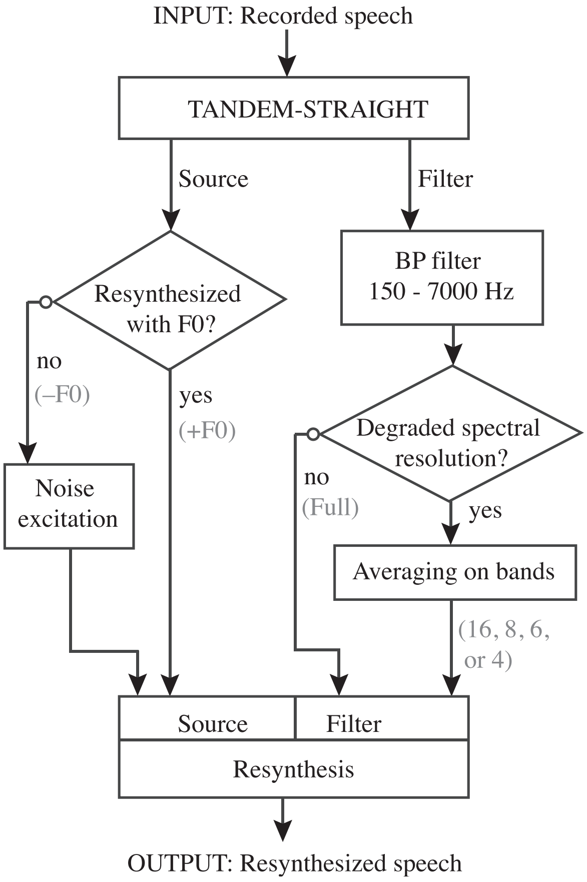

FIG. 1. Schematic of the offline processing of speech.

modified sources and filters were then recombined for resynthesis. The various conditions created with these combinations are further labeled as the number of bands (i.e., 4, 6, 8, 16, or "Full") followed by + or − F0 to indicate presence/absence of pitch. Spectrograms, spectra, and temporal envelope patterns are shown in Fig. 2(A) for the original signal (uninterrupted sentence) with only band-pass filtering (condition Full+F0), in Fig. 2(B) for the same signal when the F0 cues have been removed (condition Full−F0). Figures 2(C) and 2(D) show the spectrogram of the same signal, averaged per band for a spectral resolution of four bands with and without F0 (conditions 4+F0 and 4−F0, respectively). In these panels, the periodicity of the F0 carrier can be observed, whereas the carrier is noisy when F0 is absent. This is in line with the observation that when resynthesized in TANDEM-STRAIGHT without F0, no temporal pitch cues are left in the stimuli (Kawahara and Morise, 2011). Note that all stimuli were resynthesized with TANDEM-STRAIGHT even when no CI simulation and no pitch manipulation was applied (Full+F0, i.e., the baseline condition) to control for any possible effects of resynthesis. The sentences used in the experiment were processed under 10 voice conditions: spectral resolution {Full, 16, 8, 6, 4} × pitch {+F0, −F0}.

D. Signal processing: Interrupting speech

The specific parameters of interruptions were chosen based on a previous study (Clarke *et al.*, 2014) to avoid ceiling and floor effects in intelligibility as well as to enable a direct comparison of the results. The resynthesized sentences were interrupted online during testing by modulating with a periodic square wave of 2.2 Hz with a 50% duty cycle and a 5 ms raised cosine ramp applied on onsets and offsets to prevent spectral splatter. In the removed speech segments, interruptions were either left silent or filled with speech shaped noise (SNR of −5 dB). Filler noise bursts were produced from a single filler noise sample of 5 min duration, which was generated with white noise modulated by the long-term average spectrum of all sentences from the unprocessed (Full+F0) voice condition. The noise sample was interrupted with the inverse of the square wave used to interrupt the sentence. A 5 ms raised cosine ramp was also applied to this inverse square, so that speech and noise overlap at 50% during the transition to prevent apparent dip in the total energy. Calibration of speech was done on uninterrupted sentences, so that speech was presented at an RMS level of 65 dB sound pressure level (SPL). Noise was presented at 70 dB SPL (SNR = −5 dB).

E. Apparatus

The online processing (i.e., applying interruptions to resynthesized speech) and presentation of the stimuli were done in MATLAB on a Macintosh computer connected to an AudioFire 4 soundcard (Echo Digital Audio Corporation). The processed digital stimuli were then converted to an analog signal via a DA10 D/A converter (Lavry Engineering

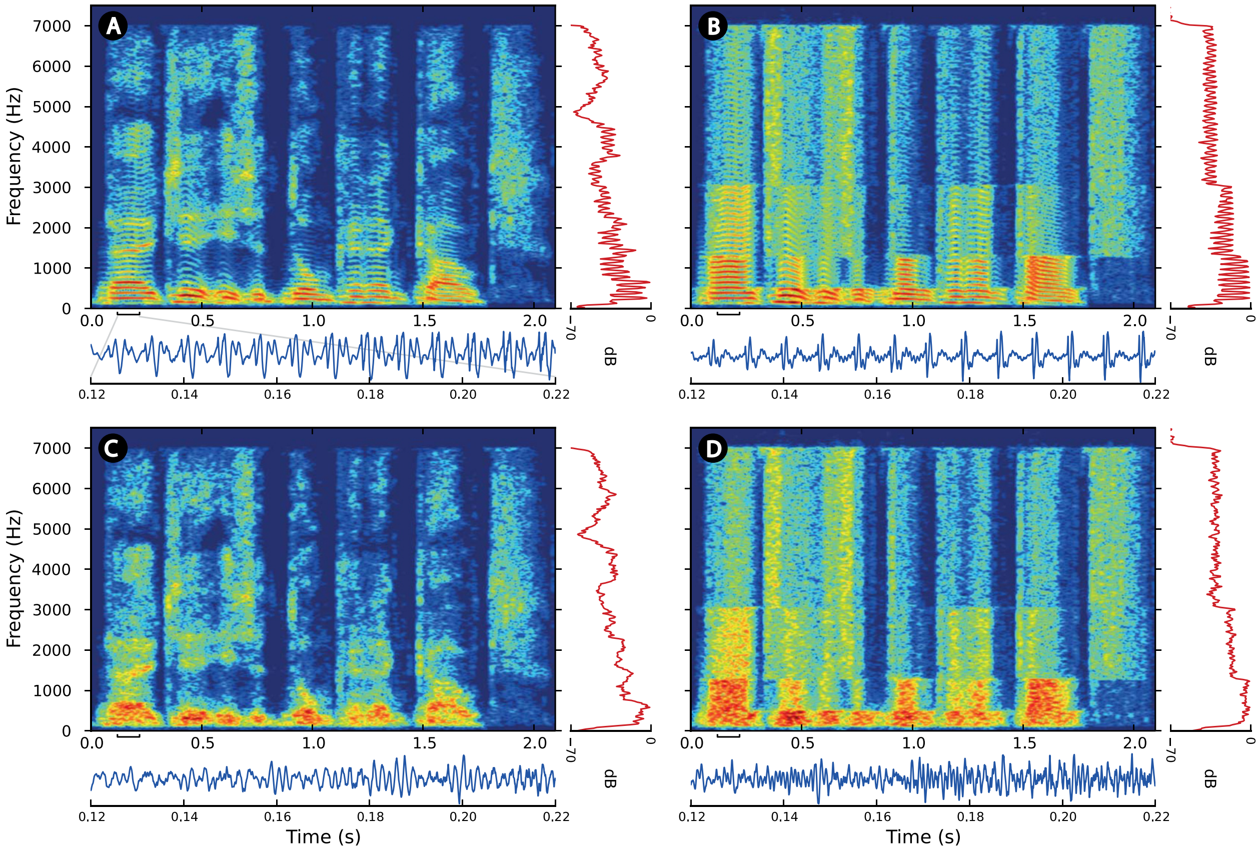

FIG. 2. (Color online) Spectrograms of the uninterrupted sentence "buiten is het donker en koud" (A) in the Full + F0 condition, (B) in the Full-F0 condition, (C) in the 4 + F0 condition, and (D) in the 4−F0 condition. Temporal envelope patterns (below each panel) and spectra (vertically on the right of each panels) are show for a 10 ms voiced segments.

Inc.) and played diotically through HD600 headphones (Sennheiser Electronic Corporation). The calibration of the stimuli was performed with a Sound and Vibration Analyser (Svan 979 from Svantek) plugged to a Kemar head (G.R.A.S.). Each participant was seated in a sound-attenuated booth. Participants' spoken responses were recorded on a PalmTrack digital voice recorder (ALESIS) for offline scoring.

F. Procedure

Participants came for a single session, which lasted around 2 h, including obtaining written informed consents, conducting the audiometric test and screening, the experimental procedure, the debriefing, and occasional breaks. The experimental procedure consisted of four parts: (1) Measuring baseline intelligibility of uninterrupted sentences with no change in spectral resolution but with or without F0 cues, (2) a short training with experimental conditions of the vocoded speech but again with uninterrupted sentences, (3) a short familiarization of actual experimental conditions with interrupted sentences, and (4) the actual data collection. In all parts of the experiment, participants were presented one sentence stimulus at a time and asked to verbally repeat what they could understand from the sentence stimulus. They were additionally encouraged to guess as much as possible. A soft, short beep preceded the stimulus to alert the listener.

1. Baseline intelligibility

Baseline intelligibility for uninterrupted speech with full spectral resolution (Full + F0 and Full−F0) was measured using the first two lists of sentences (each 13 sentences). Figure 4, solid lines, show that baseline scores decrease with decreasing spectral resolution but are not affected by the presence/absence of F0.

2. Training with vocoded speech

For familiarization with CI simulated speech, participants were trained on uninterrupted sentences in the order easier to harder voice conditions (starting from highest spectral resolution with then without F0 cues and progressing to lower spectral resolution with then without F0 cues). The eight following lists (lists 3–10) of sentences were used for this purpose. Feedback was provided to the participants: the sentence in the original voice was played followed by the CI-processed uninterrupted sentence and while its text was displayed on the screen (Benard and Başkent, 2013).

3. Familiarization with interrupted sentences

For familiarization with interrupted sentences, 4 conditions were randomly chosen from the 20 conditions used in the experiment, and sentence lists 11–15 were used (except list 13, which was previously discarded). Feedback was provided as explained in the preceding text.

4. Data collection

The experiment consisted of 20 conditions: 10 voice conditions [consisting of five spectral resolutions (Full, 16, 8, 6 and 4 bands) × two pitch conditions (with and without F0: +F0 and −F0, respectively)] × two interruption conditions (silent intervals and with noise filler). Sentence lists 16–35 were used in the experiment. The orders of the sentence lists and of the conditions were randomized.

5. Data analysis

A native Dutch speaking student assistant, who was blind to the experiment purposes, scored the recorded participant responses offline. The percent-correct scores were calculated for each sentence as the ratio of correctly identified words to the total number of words in the presented sentence. The study was designed for a repeated measures analysis of variance (ANOVA) analysis of the results, which requires homogeneity of variances. To correct for the small variances at extremes of the percentage scale, percent-correct scores were converted into RAU (rationalized arcsine units, Studebaker, 1985) to help fulfill this assumption for ANOVA. The phonemic restoration scores were calculated by subtracting the intelligibility RAU scores when the interruptions were left silent from the RAU scores when the interruptions were filled with noise. Generalized eta squared ($\eta_G{}^2$) were used to report effect sizes (Bakeman, 2005). Statistical analyses were computed in R (R Core Team, 2013).

III. RESULTS

First, the results are presented in terms of phonemic restoration effect, shown in Fig. 3. We conducted a repeated measures ANOVA on the PR scores with spectral resolution and F0 as within-subject factors. The results, summarized in Table I, showed that spectrally degrading speech reduced phonemic restoration. In contrast, on average across spectral

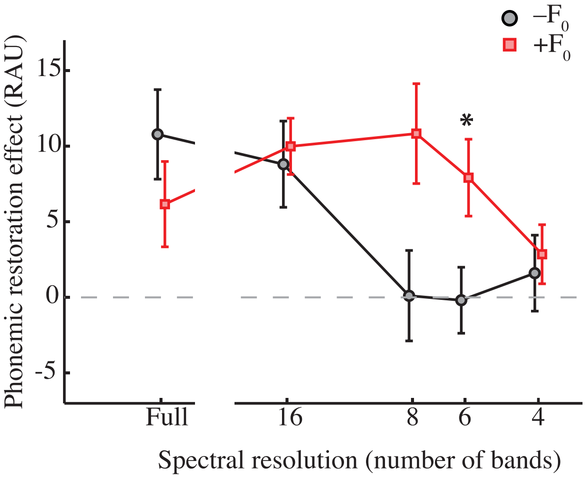

FIG. 3. (Color online) Phonemic restoration (PR) benefit as a function of spectral resolution (shown in a log scale) with (squares) or without (circles) F0. Error bars show one standard error. The star shows the significant difference between +F0 and −F0 conditions at six-bands spectral resolution.

TABLE I. Results of the two-way RM-ANOVA on PR scores. *Significant ($p < 0.05$).

Within subject factors		Effect size
SpecRes	$F_{4,72} = 3.09, p = 0.021$*	$\eta_G{}^2 = 0.055$
F0	$F_{1,18} = 2.11, p = 0.16$	$\eta_G{}^2 = 0.022$
SpecRes × F0	$F_{4,72} = 2.98, p = 0.025$*	$\eta_G{}^2 = 0.056$

resolutions, the presence or absence of F0 did not have a significant effect on the size of the phonemic restoration effect. However, a significant interaction between F0 and spectral resolution was observed showing that the effect of F0 on restoration benefit depended on the spectral resolution. More specifically, better restoration benefit was observed with the addition of F0 at a certain spectral resolution. The *post hoc* tests revealed for which spectral resolution the addition of F0 significantly increased the phonemic restoration benefit. Pairwise *t*-tests with false discovery rate (FDR) correction were used for the five relevant comparisons (at each spectral resolution). They showed that adding the F0 cue improved the restoration benefit significantly at six bands (see Table II). With the addition of F0, the improvement of restoration benefit at eight bands seemed substantial but was not significant (see Table II), although the restoration benefit became significantly different from 0 when F0 was added (see Table III). We think this lack of significance is due to noisy data that are often observed with phonemic restoration tasks, here especially because not all participants showed a consistent PR effect.

Second, we were interested in how the addition of F0 was translated into better perception of degraded speech. The results from the three-way repeated measures ANOVA are reported in Table IV, which reproduces some results from Table I on PR (our measure of interest). However, it is worth noticing that although the three main effects (interruption type, spectral resolution, and presence/absence of F0) are significant, the interaction between the spectral resolution and F0 is not. Given the interaction of these two parameters for the PR results (seen in Table IV with the three-way interaction), this effect might depend on the interruption type (SNR). Thus we considered the intelligibility scores from the interrupted speech condition (i.e., when the interruptions are left silent) as the dependent variable, shown in Fig. 4(A) (left panel). A repeated measures ANOVA was conducted on the RAU scores with F0 and spectral resolution as within-subject factors (see Table V, middle column). Both F0 and spectral resolution showed a significant main effect. First, the effect of F0 showed that the addition of F0 [square symbols in Fig. 4(A)] enhanced intelligibility, as predicted. Second, the effect of spectral resolution showed that global intelligibility of interrupted speech decreased with spectral resolution as expected. However, no interaction between F0 and spectral resolution was observed, suggesting that there was no combined effect of F0 and spectral resolution on intelligibility of interrupted sentences (note that the ANOVA on interrupted speech with noise shows similar results, see Table V, right column). *Post hoc* tests for multiple comparisons corrected with FDR were conducted on the intelligibility scores in the silent condition (see Table VI). Pairwise *t*-tests showed that intelligibility for all spectral resolution conditions significantly differed from each other, both in +F0 and −F0 conditions. Moreover, the benefit from F0 was significant for all spectral resolutions except at eight bands.

IV. DISCUSSION

The aim of the present study was to investigate independently which of reduced resolution of the spectral envelope or weaker pitch could be responsible for the smaller phonemic restoration benefit observed in actual CI users and with CI simulated speech.

A. Speech segregation

Stream segregation is necessary for speech-in-noise perception. Two different mechanisms take place for segregation: the discrimination of speech from noise and the grouping of the successive speech segments together to form a coherent stream (Bregman, 1994). In the particular case of phonemic restoration, those two mechanisms are involved also when noise is present. In the silent condition, in contrast, no discrimination between two sources is involved, thus the building up of the speech stream relies mostly on grouping. But in this case, the speech stream seems broken, which might hinder grouping and thus reduce speech intelligibility. In the filler noise condition, where both grouping and stream discrimination are involved, the filler noise can also be interpreted as masking the missing speech segments, an interpretation that the cognitive system tends to make in case of ambiguity. As a result, the speech seems more continuous and grouping of the speech segments is more efficient, thus favoring speech intelligibility. This masking interpretation, which favors grouping, would best happen when noise is similar to speech (Bashford and Warren, 1987). However, speech and noise segments must also be discriminated from each other to identify speech segments that provide the linguistic information. This discrimination would best happen when the noise is perceptually different from the speech (Dannenbring and Bregman, 1976;

TABLE II. Results of two-sided paired *t*-tests with FDR corrected *p* values for comparisons of presence or absence of pitch (+F0 vs –F0) on PR benefit at each spectral resolution. *Significant ($p < 0.05$).

Spectral resolution	Full	16 bands	8 bands	6 bands	4 bands
PR benefit	$t(18) = 1.089$ $p = 0.48$	$t(18) = 0.34$ $p = 0.74$	$t(18) = 2.046$ $p = 0.14$	$t(18) = 3.16$ $p = 0.027$*	$t(18) = 0.36$ $p = 0.74$

TABLE III. Results of two-sided paired t-tests with FDR corrected p values for comparisons of each PR benefit compared to 0. *Significant ($p < 0.001$). **Significant ($p < 0.05$). ***Significant ($p < 0.01$).

Spectral resolution	Full	16 bands	8 bands	6 bands	4 bands
+F0	$t(18) = 2.19$ $p = 0.071$	$t(18) = 5.38$ $p < 0.001$*	$t(18) = 3.28$ $p = 0.013$**	$t(18) = 3.11$ $p = 0.013$**	$t(18) = 1.47$ $p = 0.23$
–F0	$t(18) = 3.63$ $p = 0.0096$***	$t(18) = 3.093$ $p = 0.013$**	$t(18) = 0.037$ $p = 0.97$	$t(18) = 0.088$ $p = 0.97$	$t(18) = 0.64$ $p = 0.66$

Gaudrain and Carlyon, 2013; Moore and Gockel, 2002). Thus degraded spectral resolution or reduced fidelity of F0 cues (which would make the speech more noise-like) would both play against discrimination of the speech stream from the noise, the first mechanism of speech segregation, but would favor grouping more noise-like speech segments across noise, the second mechanism of speech segregation. There seems to be a balance between the two mechanisms to perform stream segregation. Given the potentially complex interaction of these parameters, in the next two paragraphs, we discuss each parameter, pitch, and spectral resolution and their effect on the discrimination and grouping mechanisms.

Regarding the effect of pitch on segregation, Clarke *et al*. (2014) showed that changing the average value of F0 across speech segments did not influence the restoration benefit. We can thereby suggest from this result that changing the average value of F0 across speech segments did not seem to influence the grouping of speech segments across the noise bursts. This was observed even when the average F0 was drastically altered, such that it changed by an octave, from that of a man to that of a woman across alternating speech segments, thus similarity of successive speech segments was violated. However, the presence of pitch itself (without considering its value) is a very strong cue that helps discriminate the noise from the speech stream. By testing the effect of presence versus absence of pitch in the present study, we expected to see an effect of F0 on segregation and therefore on restoration benefit. Consistent with our hypothesis, the phonemic restoration effect did improve with the addition of F0 when spectral resolution was degraded. However, this effect was significant only at one reduced spectral resolution condition, namely, six bands. Contrary to our expectation, adding F0 did not improve phonemic restoration at our highest (16 bands) and lowest (4 bands) degraded spectral resolution conditions.

TABLE IV. Results of the three-way RM-ANOVA on intelligibility scores. *Significant ($p < 0.001$). **Significant ($p < 0.05$).

Within subject factors		Effect size
SNR	$F_{1,18} = 45.84, p < 0.001$*	$\eta_G{}^2 = 0.084$
SpecRes	$F_{4,72} = 350.1, p < 0.001$*	$\eta_G{}^2 = 0.79$
F0	$F_{1,18} = 80.08, p < 0.001$*	$\eta_G{}^2 = 0.15$
SNR × SpecRes	$F_{4,72} = 3.086, p = 0.021$**	$\eta_G{}^2 = 0.019$
SNR × F0	$F_{1,18} = 2.11, p = 0.16$	$\eta_G{}^2 = 0.007$
SpecRes × F0	$F_{4,72} = 1.90, p = 0.12$	$\eta_G{}^2 = 0.010$
SNR × SpecRes × F0	$F_{4,72} = 2.98, p = 0.025$**	$\eta_G{}^2 = 0.019$

Note that as pitch is the percept of the F0 cues in the speech signal, timbre is related to the percept of spectral cues that can also be involved in sequential grouping, using the harmonic similarity between successive speech segments (Cusack and Roberts, 2000; Singh, 1987). Regarding the effect of spectral resolution on segregation, the significant interaction between spectral resolution and F0 found for our phonemic restoration results could be due to the fact that speech segregation can still be performed when F0 is missing, which is more likely to be so at high spectral resolution. Indeed, supporting this, when there was no degradation (i.e., at full spectral resolution), the addition of F0 did not provide an advantage for restoration as expected. At full spectral resolution, appropriate segregation was likely achieved either with or without F0. From the lack of effect of F0 at full spectral resolution, we can speculate that the spectrum of the unvoiced speech (Full−F0 condition) seems to contrast enough with that of the noise to provide a reliable discrimination cue. This is presumably also what could have happened for the high spectral resolution (16 bands). However,

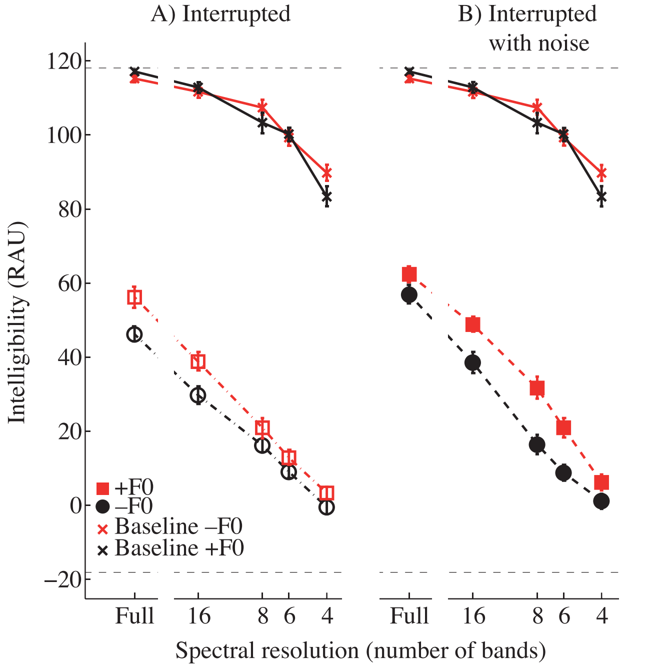

FIG. 4. (Color online) Mean intelligibility scores (in RAU) with F0 (red squares) and without F0 (black circles) as a function of spectral resolution (in log scale), (A) when interruptions are left silent (dotted lines, left panel) and (B) when interruptions are filled with noise (dashed lines, right panel). Baseline of uninterrupted sentences (repeated on both panels) are shown by the cross symbols and the solid lines. Error bars show 1 standard error.

TABLE V. Results of the two-way RM-ANOVA on interrupted speech with silent. *Significant ($p < 0.001$). **Significant ($p < 0.05$).

Within subject factors	Silent interruptions	Noise interruptions
F0	$F_{1,18} = 15.55, p < 0.001$* $\eta_G{}^2 = 0.10$	$F_{1,18} = 56.34, p < 0.001$* $\eta_G{}^2 = 0.10$
SpecRes	$F_{4,72} = 220.9, p < 0.001$* $\eta_G{}^2 = 0.78$	$F_{4,72} = 223, p < 0.001$* $\eta_G{}^2 = 0.78$
SpecRes × F0	$F_{4,72} = 1.54, p = 0.020$** $\eta_G{}^2 = 0.020$	$F_{4,72} = 3.50, p = 0.012$** $\eta_G{}^2 = 0.020$

at poor spectral resolution (four bands), there was no effect of F0 on restoration, which could mean that the spectral envelope itself is not sufficient to discriminate the speech and the noise. To further elaborate this speculation, we compared the auditory excitation patterns (computed with AIM-mat, Bleeck *et al.*, 2004) of the different speech conditions with the auditory excitation pattern of the filler noise (details of the methods can be found in Appendix A). We especially focused on the difference of the estimated perceptual distance between 16- and 4-band spectral resolution with and without F0 (details of the results can be found in Appendix B). This perceptual distance reflects the difference in auditory excitation between speech and noise. At 16-band spectral resolution, restoration benefit suggests that segregation of speech and noise still happens even when F0 cues are absent and even if the perceptual distance between speech and noise is significantly lower when F0 cues are absent from speech. Moreover, the perceptual distance between speech and noise at 4 + F0 is significantly bigger compared to 16-F0 (where it was argued previously that segregation happens). Thus segregation of speech and noise should also be possible at 4 + F0 condition. The results from the AIM output do not support that the lack of F0 benefit on phonemic restoration at four-band spectral resolution is due to the failure of discrimination between speech and noise but rather that the very poor intelligibility of the speech segments might not provide enough information to trigger a restoration benefit. This can be investigated by looking at intelligibility per se (instead of restoration effect).

B. Available speech features in interrupted speech segments

The spectral resolution had a large effect on overall intelligibility (−49 RAU from full to four-bands spectral resolution), much stronger than the presence/absence of F0 (−6 RAU from + F0 to −F0), as the effect sizes showed ($\eta_G{}^2 = 0.78$ and $\eta_G{}^2 = 0.10$ for spectral resolution and F0, respectively). Bhargava *et al.* (2014) argued that the “right kind” of bottom-up speech features in the remaining segments are needed to trigger the use of context-activated knowledge, which would benefit phonemic restoration. It is also worth highlighting that the restoration effect is not proportional to the amount of available speech information (as measured by baseline uninterrupted speech intelligibility or interrupted speech intelligibility); only that below a certain level of intelligibility, phonemic restoration seems to be unlikely to happen (Başkent, 2010) probably because the cues that are needed for restoration are missing. A similar remark can be made for a maximal level of intelligibility above which restoration would not increase further with addition of information such as F0. In this study, intelligibility (either with or without filler noise) was extremely low at four-bands spectral resolution, confirming the idea that very little speech information was available in the remaining speech segments. It is possible that with such a small amount of speech information, the speech features that would trigger phonemic restoration were not present, and even adding F0 did not provide sufficient information for restoration to happen. On the other end of the spectral resolution range, from 16 bands to full resolution, top-down restoration was already strong, and extra information could not provide any further improvement.

Regarding the effect of F0 on intelligibility, it had been shown that the combination of electric and acoustic stimulation (EAS) improves speech intelligibility (Kong *et al.*, 2005; Luo and Fu, 2006; Rader *et al.*, 2013; Turner *et al.*, 2004) compared to electric stimulation alone (CIs). This improvement was also shown for simulation of EAS compared to simulation of CIs, where vocoded speech simulate the latter and addition of low-pass filtered speech simulate the former. Başkent (2012) showed that the addition of the “acoustic” low-frequency speech to spectrally degraded speech improves overall intelligibility of interrupted speech. The non-vocoded low-frequency signal (or in the case of the implant, the acoustic stimulation) provides a number of extra cues not present in the vocoded (or electric) part, amongst which pitch is thought to be particularly important (e.g., Brown and Bacon, 2009). However, low-frequency information in EAS does not only include F0 but can also include F1 (Verschuur *et al.*, 2013). But as F0 is generally lower than F1 in average speakers, F0 is more likely than F1 to remain

TABLE VI. Results of *post hoc* pairwise *t*-tests for comparisons of intelligibility of interrupted speech (silent condition only) between spectral resolution conditions (with and without F0, first two rows, respectively) and between voicing (at each spectral resolution, last row). *Significant ($p < 0.001$). **Significant ($p < 0.05$). ***Significant ($p < 0.01$).

Spectral resolution	Full		16 bands		8 bands		6 bands		4 bands
+F0		$t(18) = -6.30$ $p < 0.001$*		$t(18) = -6.12$ $p < 0.001$*		$t(18) = -2.94$ $p = 0.011$**		$t(18) = -4.73$ $p < 0.001$*	
–F0		$t(18) = -6.55$ $p < 0.001$*		$t(18) = -6.01$ $p < 0.001$*		$t(18) = -3.46$ $p = 0.0036$***		$t(18) = -3.87$ $p = 0.0015$***	
F0 benefit	$t(18) = 3.37$ $p = 0.0043$***		$t(18) = 2.76$ $p = 0.014$**		$t(18) = 1.40$ $p = 0.18$		$t(18) = 2.15$ $p = 0.048$**		$t(18) = 2.80$ $p = 0.013$**

in residual hearing range and contribute to the EAS benefit. In this study, we investigated the benefit of F0 and not F1 (as may be the case in EAS) as F0 is completely separated from the spectral envelope information. Consistent with the importance of pitch, in the present study, the addition of F0 information improved overall intelligibility of interrupted speech (compare circles and squares in Fig. 4). Even at better spectral resolution, F0 still provided an advantage for intelligibility as confirmed by the *post hoc* tests that showed a significant difference between +F0 and −F0 conditions at different spectral resolution (see "F0 benefit" in Table VI). At full spectral resolution, this result is in line with whispered speech literature that shows that voiced speech is more intelligible than whisper (Irino *et al.*, 2012; Tartter, 1989, 1991; Vestergaard and Patterson, 2009). Whispered speech simulation also requires a spectral tilt as well as removing F0 cues (Irino *et al.*, 2012; Schwartz, 1970). However, discrimination performance in spectral envelope differences is similar between whispered and unvoiced words (Irino *et al.*, 2012), suggesting that the remaining cues in whisper (for prosodic cues: Heeren and Lorenzi, 2014; Tartter, 1989) are either also available for unvoiced speech recognition or do not provide a benefit for whisper recognition compared to unvoiced speech. In general, our F0 benefit results show that F0 seemed to provide additional speech features that improved intelligibility. These speech features can be voicing and/or pitch contours. First, F0 cues can help distinguish voiced from unvoiced consonants (even in the presence of co-varying voicing cues, such as loudness or duration—Peng *et al.*, 2012; Winn *et al.*, 2012). Second, the F0 variations (pitch contours) can give some prosodic cues of linguistic importance, such as indications on word segmentation, rhythm, or stressing the keywords in the sentence (although it was shown that prosody can be perceived in whisper via high frequency region: Heeren and Lorenzi, 2014). As argued before, when the intelligibility of the interrupted speech is very low, the speech features used for phonemic restoration are likely to be missing. Yet in that situation, adding the F0 information could bring back some of these speech features, which could then trigger restoration. This may be what happened at six bands, where adding F0 was enough to allow restoration to happen.

Başkent and Chatterjee (2010) showed that the addition of low-pass filtered speech at 500 Hz (which include F0) to noise-band vocoded speech induced a bigger increase of intelligibility at low spectral resolutions (4 and 8 bands) compared to high spectral resolutions (16 and 32 bands). However, in the present study, using sentences from the same corpus, we do not observe the same pattern of increase of interrupted speech intelligibility when F0 is added to spectrally degraded speech. Indeed the addition of F0 significantly increased intelligibility at full and 4-, 6-, 16-bands spectral resolution but not at 8-bands. This difference in results could be due to the fact that the present study added only the F0 and used an interruption rate (IR) of 2.2 Hz, whereas Başkent and Chatterjee (2010) added low-pass filtered speech below 500 Hz (LP500) and used an interruption rate of 1.5 Hz. On one hand, adding the original F0 (our +F0 conditions) carries less information than low-pass filtered speech (LP500 conditions from Başkent and Chatterjee, 2010). On the other hand, our faster IR corresponds to interruptions at the average syllabic rate in this corpus, which is a more strenuous condition for intelligibility of interrupted speech than a slower rate of 1.5 Hz IR (supported by a previous pilot study). Both these differences point to the lower overall speech intelligibility in our study. Moreover, at low spectral resolution, the addition of LP500 may include very efficient cues to help speech intelligibility (+27 and +16 RAU at four- and eight-bands spectral resolution, respectively, in Başkent and Chatterjee, 2010), whereas adding F0 may only give cues just sufficient to help intelligibility by a small amount (between +3 and +4 RAU from four- to eight-bands spectral resolution in the present study). At high spectral resolution, the addition of F0 only or LP500 provides the same benefit, probably because speech intelligibly is already quite good or because of redundancy in speech information (the information given by F0 might already be available through another cue). Also, note that the CI simulations were different. In the present study, stimuli were processed so as to completely remove pitch in the "−F0" conditions. In other more commonly used vocoders, as well as CI users, temporal F0 cues generally still provide a weak pitch percept (Moore and Carlyon, 2005), which is closer to what happens for actual CI users. Indeed Fuller *et al.* showed that CI users were able to use F0 cues for gender categorization (Fuller *et al.*, 2014); this suggests that some F0 cues are delivered in CIs. By investigating the total presence (our "+F0" conditions) versus total absence of pitch (our "−F0" conditions), we covered the full range of amount of pitch information useable for PR. Results from actual CIs and other vocoders (such as in Başkent and Chatterjee, 2010) would be expected to fall within these boundaries.

With actual CI users, Bhargava *et al.* (2014) observed phonemic restoration only for the best performers at 50% duty cycle (equal duration of ON and OFF speech segments as used in this study). In the present study, we found that the addition of F0 played a critical role for phonemic restoration only at six bands. This suggests that for CI users who do not usually show restoration, improved pitch perception could provide the sufficient extra cues to yield restoration and thus could improve their speech intelligibility in noisy environments.

V. CONCLUSION

- The combination of low spectral resolution and weak pitch representation may contribute to the poor top-down restoration of speech observed in CI simulations.
- Interaction between degraded but complementary cues seems to increase speech redundancy that helps intelligibility.
- Integrating complementary bottom-up cues (spectral resolution and pitch representation), even degraded, can help top-down restoration of speech.

ACKNOWLEDGMENTS

The authors would like to thank Hideki Kawahara for his valuable assistance in technical aspects of the work and Floor Burgerhof for transcribing participant responses as well as the participants. This study was supported by a VIDI grant from the Netherlands Organization for Scientific Research, NWO, awarded to Deniz Başkent (Grant No. 016.096.397; from Netherlands Organization for Health Research and Development, ZonMw). Further support came from a Rosalind Franklin Fellowship from University of Groningen, University Medical Center Groningen, and funds from Heinsius Houbolt Foundation. There is no conflict of interest regarding this manuscript.

APPENDIX A: METHOD FOR MODEL

The Auditory Image Model (AIM) is a functional model for human peripheral hearing. An auditory image (AI) is the initial mental representation of a sound through the peripheral auditory system before top-down processes are involved (such as attention, knowledge, and context). With a model of an average "normal-hearing" ear, we can argue that the AI represents the "objective perception" of a sound. Then comparing the AIs from different sounds would give information on how different or similar the sounds are at an early perceptual stage.[1]

The model first performs spectral and temporal analyses of incoming sounds, mimicking those that take place in the peripheral auditory system. This first stage delivers the tonotopic representation of sounds done in the cochlea. Then the model performs the channel-by-channel time-interval analyses on the neural activity pattern (NAP) that happen in the mid-brain. The last two modules of the model integrate periodicity and give a stabilized representation of NAPs (similar to the autocorrelation model of pitch perception by Meddis and O'Mard, 1997).

AIM-mat (Bleeck *et al.*, 2004) is the implementation of this model in MATLAB. In this study, it was used with the following modules: "gm2002"′ for the pre-cochlear processing, "pzfc" for the basilar membrane motion filterbank (Lyon, 2011), "hcl" for the neural activity pattern, "sf2003" for the strobe-finding, and "ti2003" for the stabilized auditory image.

APPENDIX B: RESULTS OF COMPARING AIM OUTPUTS

To compare how different or similar different sound streams are, we estimated the perceptual distance between them by calculating the Euclidean distance between the stabilized auditory images of the compared streams, i.e., the noise masker and the speech stimuli for each condition (spectral resolution and pitch). We expect that the more similar speech and noise are (i.e., the smaller the Euclidean distance is), the more masking power noise has over speech, thus the easier it is to link successive speech segments across noise, which would help intelligibility. Inversely, we expect that the more different speech and noise are (i.e., the larger the Euclidean distance is), the lesser ambiguity between the two signals there is, thus the easier it is to separate speech from noise, which would also help intelligibility.

Results of the RM-ANOVA show a main effect of spectral resolution and of F0 but also an interaction between the two parameters (presence/absence of F0 across all spectral resolutions).[1]

The *post hoc* we were most interested in showed that the Euclidean distance between the excitation patterns of noise and speech significantly increased by 235 points (6.4%) on average from 4- to 16-bands spectral resolution, $t(27) = -7.54$, $p < 0.001$, and $t(27) = -14.38$, $p < 0.001$, for + F0 and −F0, respectively. This confirms that speech at a higher spectral resolution may be coded differently from noise in the auditory nervous system. This perceptual distance is bigger than when speech is at a lower spectral resolution. Moreover, the Euclidean distance increased by 518 points (15%) with the addition of F0 at four-bands spectral resolution, $t(27) = 8.31$, $p < 0.001$. This supports the idea that speech with F0 cues may be coded differently from noise in the auditory nervous system. And this perceptual distance is bigger than when speech is unvoiced (without the F0 cues). In this case, F0 might provide a discrimination cue but the remaining segments of speech, although properly grouped in a stream, may not have provided enough information to induce phonemic restoration. This is further supported by the significantly bigger perceptual distance between speech and noise at 4+F0 compared to 16-F0, where segregation is supposed to happen as restoration benefit was observed [paired *t*-test: $t(27) = 2.090$, $p = 0.046$]. This latter result argues against the failure of discrimination between speech and noise at 4 + F0 condition.

In the interaction between spectral resolution and F0, we are particularly interested in the interaction between +F0 and −F0 at 16- and 4-band conditions. The interaction between F0 and spectral resolution is confirmed by a significantly bigger distance in the excitation patterns with the addition of F0 cues at 4- than at 16-band [paired *t*-test: $t(27) = 3.27$, $p = 0.0029$]. This indicates that adding F0 cues at 4 band should contribute more to the segregation of speech and noise than at 16-band. However, we did not observe a bigger restoration benefit when adding the F0 cues at 16- or 4-band, which suggests that the speech signal is too degraded spectrally at 4-band that even the addition of F0 cues do not provide enough bottom-up cues to trigger the top-down mechanisms of phonemic restoration.

[1] See supplementary material at http://dx.doi.org/10.1121/1.4939962 for the output of the AIM on our four selected stimuli (Figure 1) and Table I.

Adank, P., van Hout, R., and Smits, R. (**2004**). "An acoustic description of the vowels of Northern and Southern Standard Dutch," J. Acoust. Soc. Am. **116**, 1729–1738.

Bakeman, R. (**2005**). "Recommended effect size statistics for repeated measures designs," Behav. Res. Methods **37**, 379–384.

Bashford, J. A., Riener, K. R., and Warren, R. M. (**1992**). "Increasing the intelligibility of speech through multiple phonemic restorations," Percept. Psychophys. **51**, 211–217.

Bashford, J. A., and Warren, R. M. (**1987**). "Effects of spectral alternation on the intelligibility of words and sentences," Percept. Psychophys. **42**, 431–438.

Başkent, D. (**2010**). "Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores," J. Acoust. Soc. Am. **128**, EL169–EL174.

Başkent, D. (**2012**). "Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electric–acoustic stimulation," J. Assoc. Res. Otolaryngol. **13**, 683–692.

Başkent, D., and Chatterjee, M. (**2010**). "Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech," Hear. Res. **270**, 127–133.

Benard, M. R., and Başkent, D. (**2013**). "Perceptual learning of interrupted speech," PLoS One **8**, e58149.

Bhargava, P., Gaudrain, E., and Başkent, D. (**2014**). "Top-down restoration of speech in cochlear-implant users," Hear. Res. **309**, 113–123.

Bleeck, S., Ives, T., and Patterson, R. D. (**2004**). "Aim-mat: The auditory image model in MATLAB," Acta Acust. Acust. **90**(4), 781–787.

Bregman, A. S. (**1994**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (The MIT Press, Cambridge, MA), pp. 48–211.

Brown, C. A., and Bacon, S. P. (**2009**). "Low-frequency speech cues and simulated electric-acoustic hearing," J. Acoust. Soc. Am. **125**, 1658–1665.

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?," J. Acoust. Soc. Am. **95**, 3541–3554.

Clarke, J., Gaudrain, E., Chatterjee, M., and Başkent, D. (**2014**). "T'ain't the way you say it, it's what you say—perceptual continuity of voice and top-down restoration of speech," Hear. Res. **315**, 80–87.

Cusack, R., and Roberts, B. (**2000**). "Effects of differences in timbre on sequential grouping," Percept. Psychophys. **62**, 1112–1120.

Dannenbring, G. L., and Bregman, A. S. (**1976**). "Stream segregation and the illusion of overlap," J. Exp. Psychol. Hum. Percept. Perform. **2**, 544–555.

Fu, Q.-J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., and Başkent, D. (**2014**). "Gender categorization is abnormal in cochlear implant users," J. Assoc. Res. Otolaryngol. **15**, 1037–1048.

Gaudrain, E., and Carlyon, R. P. (**2013**). "Using Zebra-speech to study sequential and simultaneous speech segregation in a cochlear-implant simulation," J. Acoust. Soc. Am. **133**, 502–518.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Heeren, W. F. L., and Lorenzi, C. (**2014**). "Perception of prosody in normal and whispered French," J. Acoust. Soc. Am. **135**, 2026–2040.

Irino, T., Aoki, Y., Kawahara, H., and Patterson, R. D. (**2012**). "Comparison of performance with voiced and whispered speech in word recognition and mean-formant-frequency discrimination," Speech Commun. **54**, 998–1013.

Kawahara, H., and Morise, M. (**2011**). "Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework," SADHANA—Acad. Proc. Eng. Sci. **36**, 713–722.

Kong, Y.-Y., Stickney, G. S., and Zeng, F.-G. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**, 1351–1361.

Luo, X., and Fu, Q.-J. (**2006**). "Contribution of low-frequency acoustic information to Chinese speech recognition in cochlear implant simulations," J. Acoust. Soc. Am. **120**, 2260–2266.

Lyon, R. F. (**2011**). "Cascades of two-pole-two-zero asymmetric resonators are good models of peripheral auditory function," J. Acoust. Soc. Am. **130**, 3893–3904.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Moore, B. C. J., and Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch Neural Coding Perception, Springer Handbook of Auditory Research*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer/Birkhäuser, New York), pp. 234–277.

Moore, B. C. J., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta Acust. Acust. **88**, 320–333.

Moore, B. C. J., and Gockel, H. E. (**2012**). "Properties of auditory stream formation," Philos. Trans. R. Soc. B Biol. Sci. **367**, 919–931.

Patterson, R. D. (**2000**). "Auditory images: How complex sounds are represented in the auditory system," Acoust. Sci. Technol. **21**, 183–190.

Peng, S.-C., Chatterjee, M., and Lu, N. (**2012**). "Acoustic cue integration in speech intonation recognition with cochlear implants," Trends Amplif. **16**, 67–82.

Plack, C. J., and White, L. J. (**2000**). "Perceived continuity and pitch perception," J. Acoust. Soc. Am. **108**, 1162–1169.

Rader, T., Fastl, H., and Baumann, U. (**2013**). "Speech perception with combined electric-acoustic stimulation and bilateral cochlear implants in a multisource noise field," Ear Hear. **34**, 324–332.

R Core Team (**2013**). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).

Rubinstein, J. T. (**2004**). "How cochlear implants encode speech," Curr. Opin. Otolaryngol. Head Neck Surg. **12**, 444–448.

Samuel, A. G. (**1981**). "The role of bottom-up confirmation in the phonemic restoration illusion," J. Exp. Psychol. Hum. Percept. Perform. **7**, 1124–1131.

Schwartz, M. F. (**1970**). "Power spectral density measurements of oral and whispered speech," J. Speech Lang. Hear. Res. **13**, 445–446.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Singh, P. G. (**1987**). "Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre?," J. Acoust. Soc. Am. **82**, 886–899.

Srinivasan, S., and Wang, D. (**2005**). "A schema-based model for phonemic restoration," Speech Commun. **45**, 63–87.

Stickney, G. S., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Tartter, V. C. (**1989**). "What's in a whisper?," J. Acoust. Soc. Am. **86**, 1678–1683.

Tartter, V. C. (**1991**). "Identifiability of vowels and speakers from whispered syllables," Percept. Psychophys. **49**, 365–372.

Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (**2004**). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," J. Acoust. Soc. Am. **115**, 1729–1735.

Verschuur, C., Boland, C., Frost, E., and Constable, J. (**2013**). "The role of first formant information in simulated electro-acoustic hearing," J. Acoust. Soc. Am. **133**, 4279–4289.

Verschuure, J., and Brocaar, M. P. (**1983**). "Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise," Percept. Psychophys. **33**, 232–240.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.

Vestergaard, M. D., and Patterson, R. D. (**2009**). "Effects of voicing in the recognition of concurrent syllables (L)," J. Acoust. Soc. Am. **126**, 2860–2863.

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

Warren, R. M., and Obusek, C. J. (**1971**). "Speech perception and phonemic restorations," Percept. Psychophys. **9**, 358–362.

Winn, M. B., Chatterjee, M., and Idsardi, W. J. (**2012**). "The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing," J. Acoust. Soc. Am. **131**, 1465–1479.