NIH Public Access
Author Manuscript
Hear Res. Author manuscript; available in PMC 2012 December 1.

Published in final edited form as:
Hear Res. 2011 December ; 282(1-2): 252–264. doi:10.1016/j.heares.2011.06.004.

A model-based analysis of the "combined-stimulation advantage"

Fabien Seldran[1,2,3,4,*], **Christophe Micheyl**[5], **Eric Truy**[1,2,3,6], **Christian Berger-Vachon**[1,2,3], **Hung Thai-Van**[1,2,3,6], and **Stéphane Gallego**[1,2,3]

Fabien Seldran: fseldran@yahoo.fr; Christophe Micheyl: cmicheyl@umn.edu; Eric Truy: eric.truy@chu-lyon.fr; Christian Berger-Vachon: christian.berger-vachon@univ-lyon1.fr; Hung Thai-Van: hthaivan@gmail.com; Stéphane Gallego: sgallego@hotmail.fr

[1]INSERM U1028, Lyon Neuroscience Research Center, PACS Team (Speech, Audiology, Communication Health), Lyon, F-69000, France

[2]CNRS UMR5292, Lyon Neuroscience Research Center, PACS Team (Speech, Audiology, Communication Health), Lyon, F-69000, France

[3]University Lyon 1, Lyon, F-69000, France

[4]Vibrant Med-El Hearing Technology GmbH, 400 Ave. Roumanille, BP 309, Sophia-Antipolis 06906, France

[5]Department of Psychology, University of Minnesota, Minneapolis, MN 55455, USA

[6]Audiology and ENT Department, Edouard Herriot Hospital, Lyon F-69437, France

Abstract

Improvements in speech-recognition performance resulting from the addition of low-frequency information to electric (or vocoded) signals have attracted considerable interest in recent years. An important question is whether these improvements reflect a form of constructive perceptual interaction—whereby acoustic cues enhance the perception of electric or vocoded signals—or whether they can be explained without assuming any interaction. To address this question, speech-recognition performance was measured in 24 normal-hearing listeners using lowpass-filtered, vocoded, and "combined" (lowpass + vocoded) words presented either in quiet or in a realistic background (cafeteria noise), for different signal-to-noise ratios, different lowpass-filter cutoff frequencies, and different numbers of vocoder bands. The results of these measures were then compared to the predictions of three models of cue-combination, including a "probability summation" model and two Gaussian signal-detection-theory (SDT) models—one (the "independent noises" model) involving pre-combination noises, and the other (the "late noise" model) involving post-combination noise. Consistent with previous findings, speech-recognition performance with combined stimulation was significantly higher than performance with vocoded or lowpass stimuli alone, and it was also higher than predicted by the probability-summation model. The two Gaussian-SDT models could account quantitatively for the data. Moreover, a Bayesian model-comparison procedure demonstrated that, given the data, these two models were far more likely than the probability-summation model. Since these models do not involve any

© 2011 Published by Elsevier B.V.

*Corresponding author: INSERM U1028, CNRS UMR5292, Centre de Recherche en Neurosciences de Lyon, Equipe Audition, Hôpital Edouard Herriot, Pavillon U, Place d'Arsonval, F-69437 Lyon Cedex 03, France, Tel: +33.4.72.11.05.03, Fax: +33.4.72.11.05.04.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

constructive-interaction mechanism, this demonstrates that constructive interactions are not needed to explain the combined-stimulation benefits measured in this study. It will be important for future studies to investigate whether this conclusion generalizes to other test conditions, including real EAS, and to further test the assumptions of these different models of the combined-stimulation advantage.

Keywords

combined stimulation; cue-combination; speech recognition; signal detection theory

1. Introduction

A recent development in the field of cochlear implants (CIs) relates to "electro-acoustic stimulation" (EAS). EAS involves the combination of electric stimulation (via a CI) and acoustic stimulation (usually, via a hearing aid), either in the same ear (e.g., Gantz & Turner, 2003; Turner *et al.*, 2004) or in the opposite ear (e.g., Ching *et al.*, 2004; Kong *et al.*, 2005; Mok *et al.*, 2006). Several studies have demonstrated improved speech-recognition performance with EAS, compared to electric or acoustic stimulation alone (e.g., Büchner *et al.*, 2009; Cullington & Zeng, 2011; Gantz & Turner, 2003, 2004; Gantz *et al.*, 2004; Gantz *et al.*, 2006; Gfeller *et al.*, 2006; Turner *et al.*, 2004; von Ilberg *et al.*, 1999). Benefits of "combined stimulation" have also been observed in "simulated EAS" studies in normal-hearing listeners, in which CI processing was simulated using a vocoder, and residual low-frequency hearing was simulated by lowpass-filtering speech signals (e.g., Başkent & Chatterjee, 2010; Brown & Bacon, 2009a; Chang *et al.*, 2006; Chen & Loizou, 2010; Dorman *et al.*, 2005; Kong & Carlyon, 2007; Li & Loizou, 2008; Qin & Oxenham, 2006; Turner *et al.*, 2004).

Although the potential and actual benefits of combining acoustic stimulation with electric stimulation are well demonstrated, it is still not entirely clear what explains these benefits. It has been suggested that the provision of fundamental-frequency (F0) information at low frequencies plays a key role (e.g., Başkent & Chatterjee, 2010; Brown & Bacon, 2009a, b, 2010; Kong *et al.*, 2005; Qin & Oxenham, 2006; Turner *et al.*, 2004). F0 differences between talkers (e.g., female and male) provide a cue for the perceptual separation of concurrent voices (Brokx & Nooteboom, 1982), and it has been suggested that this cue is more salient at low frequencies (Carlyon, 1996; Culling & Darwin, 1993; Qin & Oxenham, 2006). Low-frequency F0 cues could help listeners track and extract a voice among other voices in a vocoded or electric mixture. According to another explanation, low-frequency cues could facilitate "glimpsing," i.e., selective listening to the target voice through "dips" in the temporal-envelope or spectrum of the masker (Kong & Carlyon, 2007; Li & Loizou, 2008).

While these various explanations continue to be actively investigated, a more basic question, which remains without a clear answer, is whether the perceptual benefits of combined stimulation can be explained simply in terms of non-interactive cue combination, or whether they necessarily require to assume the existence of synergetic interactions in the perceptual processing of low-frequency and electric (or vocoded) speech cues—so that access to the former somehow facilitates or enhances the use of the latter. To gain clarity on this important issue, Kong and Carlyon (2007) compared listeners' speech-recognition performance in combined-stimulation conditions to predictions derived (using a model described in section 2.4.1) under the hypothesis that the low-frequency and vocoded signals were identified independently, and that the identification decisions were then combined in an "additive" fashion. The results showed that listeners' performance was higher than

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

predicted by the model. Kong and Carlyon (2007) interpreted this result as evidence for "super-additive" effects (i.e., constructive interactions) in the combination of information across lowpass and vocoded stimuli. Chang *et al.* (2006) also noted that the performance of their listeners in combined-stimulation conditions was higher than predicted by adding the proportions of correct responses for the lowpass-alone and vocoded-alone stimuli. Earlier, Kong *et al.* (2005) pointed out that low-frequency acoustic signals that yielded essentially zero percent correct when presented alone, nonetheless enhanced speech-recognition performance when they were added to electric stimulation in the opposite ear of CI listeners. These results have been interpreted as evidence for the existence of significant synergetic interactions in the processing of lowpass- and vocoded stimuli, and have inspired the search for specific mechanisms (such as F0-guided segregation or glimpsing), which might explain these effects. However, two recent studies, one by Kong and Braida (2010), the other by Micheyl and Oxenham (submitted), indicate that it may not be necessary to posit interactions to explain the benefits of combined stimulation. Kong and Braida (2010) found that listeners' speech-recognition performance in real and simulated EAS conditions could be accounted for, to a large extent, using a non-interactive cue-combination model based on Gaussian signal detection theory (SDT). Consistent with this conclusion, Micheyl and Oxenham (submitted) reanalyzed the data of previous simulated-EAS studies using Gaussian-SDT models, and they found that such models could explain the performance of the listeners in those studies without the need to assume the existence of interactions between low-frequency and vocoded cues.

The aim of this study was to investigate further the hypothesis that listeners' ability to combine acoustic speech cues at low frequencies with temporal-envelope and degraded spectral cues contained in vocoded signals at higher frequencies, can be predicted using relatively simple psychophysical models of cue-combination, which do not posit any interaction in the processing of the two types of cues. To this aim, speech-recognition performance was measured using lowpass-filtered, vocoded, and combined (lowpass-filtered + vocoded) signals under a variety of conditions—including different signal-to-noise ratios (SNRs), different lowpass-filter cutoff frequencies, and different numbers of vocoder bands—in 24 normal-hearing listeners. The results of these measures were then compared to predictions obtained using three models of cue combination, including the model used by Kong and Carlyon (2007), and two Gaussian-SDT models.

2. Methods

2.1 Subjects

Twenty-four subjects (11 female, 13 male, aged 19–35 years, mean = 23.6 years) took part in the study. All had normal hearing, defined as hearing thresholds of 20 dB HL or better at octave frequencies between 250 and 8000 Hz. In accord with the Declaration of Helsinki, written informed consent was obtained from all subjects prior to their inclusion in the study.

2.2. Stimuli and procedure

The speech stimuli consisted of 40 lists of 10 disyllabic words (Fournier, 1951) spoken by a single male talker and recorded on a compact disc (CD, 44.1-kHz sampling rate, with 16-bit quantization range). The "lowpass" (L) stimuli were produced by filtering these signals digitally (using the FFT-filter function of the Adobe Audition software) below a cutoff frequency (CF) of 500, 707, 1000, or 1414 Hz (stopband attenuation > 70 dB). The vocoded (V) stimuli were produced by, firstly, bandpass-filtering the original speech signals into N = 1, 2, 3, or 4 frequency bands, which are hereafter referred to as "analysis bands." The cutoff frequencies of the analysis bands are listed in Table I. For the highest CF (1414 Hz), only two N conditions were tested: N = 1 and N = 2. The lower frequency limit of the lowest

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

analysis band was equal to CF. The upper frequency limit of the highest analysis band was fixed at 4000 Hz. The center frequencies of the analysis bands were equidistant on a logarithmic scale between CF and 4000 Hz. The temporal envelopes of the signals at the output of these analysis bands were then extracted using full-wave rectification, followed by lowpass-filtering at 50 Hz to eliminate pitch-related envelope fluctuations (Carroll & Zeng, 2007). Thirdly, the resulting temporal envelopes were used to modulate the amplitude of noise bands ("synthesis bands"). To avoid interactions between modulation sidebands, each synthesis band was 150-Hz (75-Hz on each side) narrower than the corresponding analysis band (Carroll & Zeng, 2007). Finally, the amplitude-modulated synthesis bands were scaled to have the same RMS amplitude as the corresponding analysis bands in the original signal, then they were summed to produce the final V signal. The "combined" (C) stimuli were produced by adding together L and V signals corresponding to the same CF.

Stimuli were either presented in quiet, or added to recordings of cafeteria noise. The latter were chosen to provide a realistic, everyday-life noise background. They contained unintelligible chatter from a large number of talkers, and occasional noises of clanging dishes or chairs being moved, typical of a cafeteria. The noise was processed in the same way as the speech. The processed speech and noise signals were recorded separately on the left and right tracks of a CD. During the tests, the signals from the left and right tracks were mixed at different signal-to-noise ratios (SNRs): −6 dB, 0 dB, and +6 dB. For the Quiet condition (which may be thought of as corresponding to an infinitely large physical SNR, and is hereafter included as a fourth SNR condition), the cafeteria noise was turned off.

Together, three presentation modes (L, V, and C), four CFs, and four SNRs would have resulted in a total of 48 stimulus conditions. However, because it was anticipated that performance in the quiet C conditions with N = 3 or 4 and a CF of 1000 Hz would be near ceiling (100% correct), the 1414-Hz CF was tested only when N was less than 3. As a result, the number of conditions tested was equal to 32. For each condition, and each listener, one list of 10 dissyllabic words was drawn at random (without replacement) from the 40 lists of disyllabic words. The 10 words were presented to the listener's right ear. Listeners were instructed to repeat each word after its presentation. For scoring, the total number of correctly repeated syllables in each test condition was divided by the number of syllables (20) and multiplied by 100 to yield a percent-correct (PC) score. The tests took approximately three hours per listener. Testing was divided into two sessions of approximately 90 minutes each, on different days. The listeners did not receive training prior to the tests.

2.3 Apparatus

The stimuli were played on a CD player (PHILIPS – CD723) connected to an audiometer (MADSEN – Orbiter 922), and delivered through TDH-39 earphones. Tests took place in a soundproof booth at Edouard Herriot Hospital in Lyon. The study was approved by the local Ethics Committee (CPP Sud-Est IV, Centre Leon Berard de Lyon, France, N° ID RCB: 2008-A01479-46).

2.4 Models

2.4.1. The probability-summation model—The first model that was considered in this study has been used in various other contexts in the psychophysical literature, where it is sometimes referred to as the "probability summation," or "independent decisions," model (for introductions, see: Green & Swets, 1966; Macmillan & Creelman, 2005; Treisman, 1998). The model has been used to predict the detection of simple events (e.g., signal versus no-signal) at the outputs of two or more sensory channels (e.g., Pelli, 1985; Pirenne, 1943). Fletcher (1953) used this model to predict the proportion of correct responses for the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

recognition of simultaneously presented bands of speech based on the intelligibility of each band in isolation (see also: Braida, 1991; Ronan *et al.*, 2004). Boothroyd and Nittrouer also used, and expanded, the model to analyze context effects in speech recognition (Boothroyd & Nittrouer, 1988; Nittrouer & Boothroyd, 1990). The latter work prompted the use of the model in a simulated-EAS study by Kong and Carlyon (2007). In this context, the probability of a correct response in combined-stimulation conditions is predicted as the complement of the product of the error probabilities in the corresponding non-combined stimulation conditions, as follows.

$$P_{C\_PS} = 1 - (1 - P_L)(1 - P_V). \tag{1}$$

In this equation, and the ones that follow, P_L denotes the probability of a correct response in the L condition, P_V denotes the probability of a correct response in the V condition, and P_C denotes the predicted probability of a correct response in the C condition; the subscript "_PS" was added to distinguish the predictions of the probability-summation model from the predictions of other models.

2.4.2 Gaussian-SDT models—The model described in the previous section falls in the category of "post-labeling" models (Braida, 1991; Ronan *et al.*, 2004). These models assume that listeners, first, identify speech items (phonemes, syllables, or words) within each channel, then, combine the resulting identification decisions in some way, e.g., by selecting one of the two answers with a certain probability. If the observations on which the decisions are based are not dichotomous, this decision strategy is sub-optimal, meaning that it is possible to achieve higher performance than predicted by these models, simply, by using a different decision strategy. In general, optimal (maximum-likelihood) decision strategies involve a combination of continuous decision variables, or observations, across channels *before* a decision is made as to which item was presented (Green & Swets, 1966; Wickens, 2001). Models employing this type of decision strategy are known as "pre-labeling" models (Braida, 1991; Ronan *et al.*, 2004). Usually, the pieces of information that are (assumed to be) combined in these models are continuous quantities, which are related to likelihoods or likelihood ratios—for instance, the likelihood that speech item x was presented, given that input y_i was received in channel i. Moreover, these models usually assume that listeners' ability to correctly identify speech signals is limited by variability, or "noise." The "noise" can be *external*, such as background noise, a competing voice, or variability in speech signals due to within- and across-speaker variations (Uchanski & Braida, 1998), or internal (e.g., neural noise, or fluctuations in attention over time). For tractability, and justified by the central-limit theorem (see: Green & Swets, 1966), the noise is traditionally assumed to have a Gaussian probability distribution.

Depending on whether the noise that limits performance is assumed to occur before or after the combination of information across channels, two models are obtained: the independent-noises model, and the post-combination noise model.

2.4.2.1 The independent-noises model: In this model, d' for the combined-stimulation case (hereafter denoted as d'_C) is related to d' for lowpass-filtered stimulation alone (hereafter denoted as d'_L) and to d' for vocoded stimulation alone (hereafter denoted as d'_V) by,

$$d'_{C\_IN} = \sqrt{d'^2_L + d'^2_V}. \tag{2}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The subscript, IN, stands for "independent noises" model. The PCs measured in the L and V conditions can be used to estimate d'_L and d'_V by inverting (numerically) the following integral equation, which describes the relationship between PC and d' for the m-alternative forced-choice (mAFC) identification task (see: Green & Birdsall, 1958; Green & Dai, 1991).

$$PC = \int_{-\infty}^{+\infty} \varphi(z - d')\Phi^{m-1}(z)dz. \tag{3}$$

In this equation, $\varphi(.)$ denotes the standard normal probability density function, $\Phi^{m-1}(.)$ denotes the cumulative standard normal distribution raised to the power $m-1$, where m is the number of response alternatives.

Based on a reanalysis of Kryter's (1962) data on speech intelligibility as a function of set-size, Müsch and Buus (2001) found that, when the number of items (e.g., words) in a speech-recognition test is large (e.g., one hundred or more), setting m to 8000 leads to more accurate predictions than setting it to the size of the stimulus set. A similar effect was observed by Green and Birdsall (1958), who suggested that when the stimulus set is large and its contents are not known in advance to the listener, performance is determined, not by the number of stimulus alternatives, but by the size of the listener's active vocabulary. Müsch and Buus (2001) estimated the size of this active vocabulary to be equal to 8000. Accordingly, in the analyses described below, the parameter m was initially set to 8000. However, additional analyses were also performed, in which, either the value of m was treated as a free parameter, or uncertainty regarding the value of m was explicitly taken into account, using a Bayesian approach (described in the Appendix).

Combining Eqs. 2 and 3, the prediction of the independent-noises model for the combined condition is given by,

$$P_{C\_IN} = \int_{-\infty}^{+\infty} \varphi\left(z - \sqrt{d'^2_L + d'^2_V}\right)\Phi^{m-1}(z)dz. \tag{4}$$

2.4.2.2 The late-noise model: Whereas the independent-noises model assumes that the only significant source of noise occur *before* information is combined across the L and V signals, the "late-noise" model assumes that the only significant source of performance-limiting noise occurs *after* the combination. In this model, d' for the combined (C) case is predicted using the following equation.

$$d'_{C\_LN} = d'_L + d'_V. \tag{5}$$

where the subscript, LN, stands for "late noise."

Combining Eqs. 3 and 5, the PC prediction of the late-noise model for the case of combined stimulation is obtained as,

$$P_{C\_LN} = \int_{-\infty}^{+\infty} \varphi(z - d'_L - d'_V)\Phi^{m-1}(z)dz. \tag{6}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

3. Results

3.1. PCs measures in human listeners

Figure 1 provides an overview of the data. The upper panel shows how PC varied as a function of N and SNR, with CF as a parameter, for the three stimulation modes: V, L, and C. Many of the trends that are apparent in this figure are as expected based on earlier findings. In particular PC usually increased with CF, SNR and, for the V and C stimulation modes, it also increased with N. Also consistent with previous findings, PCs measured in the C stimulation mode were generally (with few exceptions) higher than the PC measured (in corresponding SNR, CF, and when applicable, N, conditions) for the L or V stimulation modes.

The statistical significance of these observations was confirmed by the results of repeated-measures ANOVAs, as detailed in the following five sub-sections. These analyses were preceded by Mauchly's test of sphericity. Whenever the sphericity assumption was violated, the Greenhouse-Geisser correction was applied; the *F* and *p* values reported in this article include this correction, as needed. Although the results reported below were obtained using analyses performed directly on the PC data, parallel analyses were performed on d' values computed using Eq. 3 (with *m* set to 8000). Unless noted otherwise, these analyses yielded qualitatively similar outcomes and, in the interest of space, only the results of analyses performed on PC data are reported.

3.1.1 Analysis of results for the L stimulation mode—A two-way (SNR × CF) ANOVA on the PCs corresponding to the L stimulation mode showed significant main effects of SNR ($F(3, 69) = 154.90$, $p < 0.0005$) and CF ($F(3, 69) = 84.69$, $p < 0.0005$), and a significant interaction ($F(9, 207) = 4.50$, $p < 0.0005$).

3.1.2 Analysis of results for the V stimulation mode—A three-way (N × SNR × CF) ANOVA on the PCs corresponding to the V stimulation mode showed significant main effects of N ($F(3, 69) = 246.87$, $p < 0.0005$), SNR ($F(3, 69) = 83.08$, $p < 0.0005$), and CF ($F(2, 46) = 7.39$, $p = 0.002$), as well as significant interactions between N and SNR ($F(9, 207) = 60.40$, $p < 0.0005$), SNR and CF ($F(6, 138) = 8.83$, $p < 0.0005$), and a three-way interaction ($F(18, 414) = 2.71$, $p = 0.019$). The N-by-CF interaction failed to reach statistical significance ($F(6, 138) = 2.43$, $p = 0.057$).

3.1.3 Analysis of results for the C stimulation mode—A three-way (N × SNR × CF) ANOVA on the PCs corresponding to the C stimulation mode showed significant main effects of N ($F(3, 69) = 76.60$, $p < 0.0005$), SNR ($F(3, 69) = 83.08$, $p < 0.0005$), and CF ($F(2, 46) = 148.27$, $p < 0.0005$), as well as significant interactions between N and SNR ($F(9, 207) = 3.41$, $p = 0.004$), N and CF ($F(6, 138) = 12.67$, $p < 0.0005$), SNR and CF ($F(6, 138) = 3.60$, $p = 0.008$), and a three-way interaction ($F(18, 414) = 1.90$, $p = 0.048$). The same ANOVA, when applied on d', also showed significant main effects of N ($F(3, 69) = 152.890$, $p < 0.0005$), SNR ($F(3, 69) = 309.11$, $p< 0.0005$), and CF ($F(2, 46) = 146.21$, $p < 0.0005$). However, except for the N-by-CF interaction ($F(6, 138) = 9.05$, $p < 0.0005$), this analysis showed no significant two- or three-way interaction.

3.1.4 Comparisons between L and C stimulation modes—Planned contrasts analyses comparing PC measured in the L stimulation mode against the PCs measured in corresponding SNR and CF conditions of the C stimulation mode, separately for each N, showed significant differences between the two stimulation modes for all Ns ($F(1, 23) = 83.04$, $p < 0.0005$ for N = 1; $F(1, 23) = 475.36$, $p < 0.0005$ for N = 2; $F(1, 23) = 901.00$, $p < 0.0005$ for N = 3; and $F(1, 23) = 872.65$, $p < 0.0005$ for N = 4).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

3.1.5 Comparisons between V and C stimulation modes—A four-way (stimulation mode × N × SNR × CF) ANOVA on PCs measured using the V and C stimulation modes showed a significant main effect of the stimulation mode ($F(1, 23) = 3529.91$, $p < 0.0005$). Significant interactions were also observed between the stimulation mode and the following factors: N ($F(3, 69) = 24.168$, $p < 0.0005$), SNR ($F(3, 69) = 176.76$, $p < 0.0005$), CF ($F(2, 46) = 148.94$, $p < 0.0005$). Additionally, the stimulation-mode factor was involved in significant three-way interactions with N and SNR ($F(9, 207) = 17.73$, $p < 0.0005$), and N and CF ($F(6, 138) = 7.13$, $p < 0.0005$); the three-way interaction between stimulation mode, SNR, and CF failed to reach statistical significance ($F(6, 138) = 2.33$, $p = 0.057$). Finally, the four-way (stimulation mode × N × SNR × CF) interaction was significant ($F(18, 414) = 2.07$, $p = 0.035$). A parallel analysis on d' yielded qualitatively identical outcomes, except for the four-way interaction, which was not statistically significant ($F(18, 414) = 1.66$, $p = 0.111$).

3.2 Model predictions

In this section, the predictions of the above-described cue-combination models are compared to the PCs measured for the C stimulation mode. The predicted PCs were obtained by setting the variables P_L and P_V in the equations in section 2.4 to the mean PCs measured in L and V conditions in the listeners. For the SDT models, predictions could not be computed when the mean PC for the L or V presentation mode was equal to zero, as this led to an undefined d' . To circumvent this problem, mean PC values that were equal to zero were replaced by $1/m$ (i.e., the chance rate) prior to the computation of d', resulting in a d' of zero for that condition. Only five out of a total of 72 mean PC values had to be replaced in this way. Four of the replaced values corresponded to conditions involving one-band V stimuli; the remaining value corresponded to a condition involving two-band V stimuli.

3.2.1 Probability-summation model—The predictions of the probability-summation model are shown in the upper panel of Fig. 2 (empty symbols), together with the mean PCs measured in the listeners for the C stimulation mode (filled symbols). The differences between the data and the predictions are shown in the lower panel. These differences were obtained by subtracting the predicted PCs from the observed PCs, so that positive values indicate that listeners' performance exceeded the predictions of the model. With a few exceptions, which occurred in conditions where N was equal to 1, the mean PCs of the human listeners were higher than predicted by the probability-summation model. In most conditions, the discrepancy between data and predictions was larger than 10 percentage points. In some conditions—especially, when N was equal to 4—the discrepancies were as large as 40 to 50 percentage points. For this model, the RMS difference between data and predictions across all N, SNR, and CF conditions was equal to 24 percentage points.

3.2.2 Independent-noises model—Figure 3 shows the predictions of the independent-noises model with the parameter m set to 8000. Although these predictions were almost always lower than the mean PCs of the listeners, the deviations between data and predictions were generally smaller than for the probability-summation model (compare the lower panels of Figs. 2 and 3). The RMS difference between data and predictions was equal to 13 percentage points. This approximately half the RMS error for the probability-summation model (24 percentage points). However, note that, for many of the conditions tested, the predictions of the independent-noises model were more than 10 percentage-points lower than the mean PCs of the human listeners.

As mentioned above, the predictions in Fig. 3 were obtained with the parameter, m, set to 8000. As explained in section 2.2, this choice was based on the conclusion of an earlier study, which estimated the size of listeners' active vocabulary in open-set speech tests at

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

about 8000 words (Müsch & Buus, 2001). However, it is unclear whether this conclusion applies to the conditions and listeners tested in the current study. To explore the possibility that the independent-noises model could fit the listeners' data more closely for different settings of the parameter *m*, a gradient-descent algorithm was used to find the value of *m* that minimized the RMS error between data and predictions. The results, which are shown in Fig. 4, reveal that when *m* was adjusted in this way, the independent-noises model could fit the human-listeners' data with an error of less than 10 percentage points in most of the conditions tested, and an overall RMS error of only 7 percentage points. For comparison, the RMS deviation between the listeners' mean PCs and the upper (or lower) bounds of the 95%-confidence intervals around these mean PCs was equal to 6 percentage points; thus, the RMS error between data and predictions was not appreciably larger than the variability of the data. However, it is worth noting that the *m* value that was found to yield such good agreement between data and predictions was considerably larger than the estimate obtained by Müsch and Buus (2001), being equal to 870000. Possible explanations for this result are considered in the discussion.

3.2.3 Late-noise model—The predictions of the late-noise model with the parameter *m* set to 8000 are shown in Fig. 5. In contrast to the probability-summation and independent-observations models, this model almost invariably *over*-predicted listeners' performance. The RMS error between data and predictions was equal to 18 percentage points.

Fig. 6 shows the predictions that were obtained for the late-noise model when the value of the parameter, *m*, was adjusted to minimize the RMS error between the data and the predictions. The value of *m* that was found to minimize the RMS error was equal to 653. For this setting of *m*, the RMS error between the predictions of the model and the data was equal to 11 percentage points.

3.3. Bayesian model comparison

The finding that smaller RMS errors could be obtained for the two Gaussian-SDT models than for the probability-summation model was not entirely unexpected, since the former two contain one free parameter, whereas the latter contains none. To compare the ability of these models to "explain" the data (in a statistical sense), while taking into account the different numbers of free parameters, we used a Bayesian model-comparison procedure. The mathematical details of this procedure are described in the Appendix. One advantage of the Bayesian approach is that "Occam's razor" principle, which penalizes models for having more degrees of freedom, is implemented naturally and automatically via the marginalization of the likelihood over the space of model parameters (see: MacKay, 2003, chapter 28, pp. 343–356). Other advantageous features of the Bayesian approach to model-based analyses of psychophysical data are explained in Rouder and Lu (2005). One such advantage, which was especially important in the context of the current study, is that Bayesian inference provides a theoretically principled way to deal with correct-response counts equal to zero. This was especially important here, because the predictions of the models considered in this study are based upon estimates of probabilities of correct-response in V conditions, in which correct-response counts of zero were sometimes observed. With relatively small numbers of trials per condition, it is possible to obtain correct-response counts of zero, even if the underlying probability of a correct response is higher than the rate chance ($1/m$). The Bayesian model-comparison approach described in the Appendix takes this into account.

The outcome of a Bayesian model-comparison analysis is traditionally summarized using Bayes factors. Bayes factors indicate the relative likelihood of a model, compared to another model, given the data. Figure 7 shows Bayes factors for the independent-noises model,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

compared to the probability-summation model. The values shown in this figure were obtained by taking the geometric mean, across all listeners, of the "individual" Bayes factors that were computed (using the approach described in the Appendix) for each combination of CF, N, and SNR. Values higher than 1 (indicated by the horizontal line) mean that, for the considered condition, the independent-noises model was more likely than the probability-summation model, given the data; values lower than 1 indicate the opposite. For example, a value of 10 means that, on average across all listeners, the independent-noises model was 10 times more likely to have generated the individual data than the probability-summation model. As can be seen, except for some of the lowest-SNR, lowest-CFs, and lowest-N conditions, the mean Bayes factors in Fig. 7 were generally higher than 1. On average across all conditions and listeners, the mean Bayes factor was 41.73, indicating that the independent-noises model was about 42 times more likely to have generated the individual data than the probability-summation model. A similar comparison between the late-noise model and the probability-summation model yielded a mean Bayes factor of 26.08. The mean Bayes factor for the comparison between the independent-noises model and the late-noise models was smaller, being equal to 1.60.

In interpreting these values and those shown in Fig. 7, it is important to note that they are means of Bayes factors computed based on individual data. The "group-level" Bayes factors, which were computed by taking into account the data of all listeners at once (as explained in the Appendix), were several orders of magnitude larger than these "individual" Bayes factors. The geometric-mean (across all conditions) of the group-level Bayes factor for the independent-noises model compared to the probability-summation model was equal to 5.85E+38. The geometric-mean group-level Bayes factor for the late-noise model compared to the probability-summation model was equal to 1.69E+33. The geometric-mean group-level Bayes factor for the independent-noise model compared to the late-noise model was equal to 3.47E+05. According to Jeffreys' scale (Jeffreys, 1961), Bayes factors of 6 or more provide "strong evidence" for one model over another, and Bayes factor larger than 100 provide "conclusive evidence."

4. Discussion

Consistent with previous findings (e.g., Chang *et al.*, 2006; Kong & Carlyon, 2007; Li & Loizou, 2008; Qin & Oxenham, 2006), the results of this study showed generally higher speech-recognition performance with combined (lowpass-filtered + vocoded) stimulation than with lowpass-filtered or vocoded stimuli alone. This "combined-stimulation advantage" was observed under a wide variety of conditions, including in quiet and in a realistic noise background (cafeteria), for different SNRs (ranging from −6 to +6 dB), and different CFs (ranging from 500 to 1414 Hz). The numbers of vocoder bands (N) that were tested in this study were relatively small, ranging from 1 to 4. Interestingly, even with so few vocoder bands, speech-recognition performance was generally higher when both vocoded and lowpass-filtered signals were presented, than when only lowpass-filtered signals were presented. These findings confirm and extend earlier demonstrations of combined-stimulation advantages under simulated EAS conditions in normal-hearing listeners.

The results of the current study are also consistent with the findings of Kong and Carlyon (2007), who were the first to test the "super-additivity" hypothesis rigorously, using a mathematical model. These authors observed that their listeners' performance in combined-stimulation (lowpass-filtered + vocoded) conditions was higher than predicted by the probability-summation model. A similar outcome was obtained in the current study. However, as noted above, the probability-summation model is largely suboptimal. The two Gaussian-SDT models that were evaluated in this study both predicted higher performance in combined-stimulation conditions than the probability-summation model, and even when

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

taking into account the fact that these models included a free parameter (while the probability-summation model did not), they were found to provide a far more plausible account of the human-listeners' data than the probability-summation model. This outcome suggests that future efforts to develop predictive models of human listeners' speech-recognition performance in combined-stimulation conditions should build upon Gaussian SDT models, rather than upon the probability-summation model.

When the parameter, *m*, in the two Gaussian-SDT models was adjusted to minimize the error between the predictions and the data, both models were found to be able to fit the human-listeners' speech-recognition performance in combined-stimulation conditions well. The RMS error between data and predictions (across all tested conditions) was equal to 7 percentage points for the independent-noises model, and to 11 percentage points for the late-noise model. Since neither of these models involves any interaction in the combination of information, this indicates that, for the listeners and stimulus conditions tested in this study, the "combined-stimulation advantage" can be explained without invoking "constructive interactions" in the perceptual processing of the low-frequency and vocoded signals. We emphasize that this result does not invalidate explanations of the combined-stimulation advantage that imply a form of constructive interaction. However, it indicates that such explanations may not always be needed. A similar conclusion was reached in a recent study by Kong and Braida (2010). These authors concluded that their listeners' speech-recognition performance in simulated EAS conditions could be accounted for using "pre-labeling" models, which were based on Gaussian-SDT assumptions. As those evaluated in the current study, these SDT models did not assume any interaction in the processing of lowpass-filtered and vocoded signals, but they did assume optimal (or quasi-optimal) combination of the cues contained in these signals. More recently, Micheyl and Oxenham (submitted) found that Gaussian-SDT models could account for the data of Kong and Carlyon (2007). Nonetheless, it is important to acknowledge that the use of a syllable identification task with single-word stimuli, and of a noise vocoder with a 50-Hz cutoff frequency for the envelope lowpass filter (which probably removed even the weakest pitch cues), may have created an experimental environment that was perhaps more conducive to simple additive effects than others. Additional work is needed to further clarify the conditions under which the observed benefits of combined stimulation can, or cannot, be accounted for without positing some form of interaction. Model-based analyses such as those described here and in other articles (e.g., Braida, 1991; Kong & Braida, 2010; Kong & Carlyon, 2007; Micheyl & Oxenham, submitted; Müsch & Buus, 2001; Ronan *et al.*, 2004), should certainly play a key role in this endeavor.

Two issues with the conclusion that the two Gaussian-SDT models tested in this study could explain the data deserve mention. First, the value of *m* that was needed to minimize the RMS error between the predictions of the independent-noises model and the listeners' data was quite large, being equal to 870000. This is more than two orders of magnitude larger than the estimate of Müsch and Buus (2001), and it appears to be unrealistically large for an estimate of the size of listeners' active vocabulary—although it is important to note that the nature and number of speech "templates" that can be stored in the human brain, and that may be needed to make speech recognition possible under a wide variety of acoustic conditions, are currently unknown. This limits the appeal of the independent-noises model. Although the *m* value needed to minimize the RMS error between data and predictions for the late-noise model was considerably smaller (being equal to 653), the assumption on which this model is based, according to which the predominant source of performance-limiting noise occurs after low-frequency and vocoded cues are combined, also seems unrealistic. In low-SNR conditions, where relatively high levels of external noise are present, the noise that limits performance is likely introduced *before*, rather than *after*, information is combined across channels. This limits the appeal of the late-noise model

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

somewhat. However, note that these limitations can be overcome, simply, by combining pre- and post-combination noise sources within the same model, and by letting the magnitude of the pre-combination noise decrease as the SNR increases. A possibility, which we did not consider in the current study, but which it would be interesting to investigate in future experiments, is that increasing external noise increases the uncertainty of the listener, and forces the listener to search through a larger set of templates (which may be thought of as an increase in the listeners' active vocabulary).

Two important limitations of this study must be pointed out. Firstly, we only considered models that did not involve interactions in the perceptual processing of low-frequency and vocoded cues. Our decision on this point was in accord with the stated goal of this study: to test whether the EAS advantage could be accounted for quantitatively without invoking interactions. The decision can be further justified *a posteriori*, given that the answer to this question was found to be positive—thus making the need to search for alternative (and less parsimonious) explanations less pressing. However, in future work, it would be interesting to investigate models of the EAS advantage that involve across-channel interactions. Moreover, it would be important to examine the impact of correlations (e.g, temporal-envelope correlations) or, more generally, statistical dependencies, between low-frequency and vocoded (or electric) speech signals. Secondly, and finally, an important caveat, which applies to all simulated-EAS studies, relates to use of noise-vocoding and lowpass-filtering in normal-hearing listeners to simulate CI processing and residual low-frequency hearing in hearing-impaired listeners. Important aspects and consequences of cochlear damages, such as reduced frequency selectivity at low frequencies (Faulkner *et al.*, 1990; Moore, 1985; Moore *et al.*, 1997; Tyler, 1986) and injury-induced neural degeneration (Kujawa & Liberman, 2009), which can greatly limit the benefit of combined acoustic and electric stimulation in hearing-impaired individuals, are not taken into account in these simulations. Therefore, while EAS simulations in normal-hearing listeners provide a test-bed for the development and the evaluation of explanatory or predictive models of the EAS advantage, studies in CI listeners are crucially needed to validate these models.

Acknowledgments

This work was supported by Vibrant Med-El France (Doctoral Research Grant CIFRE 266/2007 to F.S.), the French National Center for Scientific Research (CNRS), NIH R01 DC05216 (author C.M.) and "Laboratoire Audition Conseil" (22 rue Constantine, Lyon 69001, France).

Abbreviations

CF	cutoff frequency
CI	cochlear implant
EAS	electro-acoustic stimulation
F0	fundamental frequency
PC	percentage of correct responses
RMS	root-mean-square
SDT	signal detection theory
SNR	signal-to-noise ratio

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

References

Başkent D, Chatterjee M. Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech. Hear Res. 2010; 270:127–133. [PubMed: 20817081]

Bishop, C. Pattern Recognition and Machine Learning. Springer; Berlin: 2006.

Boothroyd A, Nittrouer S. Mathematical treatment of context effects in phoneme and word recognition. J Acoust Soc Am. 1988; 84:101–14. [PubMed: 3411038]

Braida LD. Crossmodal integration in the identification of consonant segments. Q J Exp Psychol A. 1991; 43:647–77. [PubMed: 1775661]

Brokx JP, Nooteboom SG. Intonation and the perceptual separation of simultaneous voices. J Phonetics. 1982; 10:23–36.

Brown CA, Bacon SP. Low-frequency speech cues and simulated electric-acoustic hearing. J Acoust Soc Am. 2009a; 125:1658–65. [PubMed: 19275323]

Brown CA, Bacon SP. Achieving electric-acoustic benefit with a modulated tone. Ear Hear. 2009b; 30:489–93. [PubMed: 19546806]

Brown CA, Bacon SP. Fundamental frequency and speech intelligibility in background noise. Hear Res. 2010; 266:52–9. [PubMed: 19748564]

Büchner A, Schüssler M, Battmer RD, Stover T, Lesinski-Schiedat A, Lenarz T. Impact of low-frequency hearing. Audiol Neurootol. 2009; 14(Suppl 1):8–13. [PubMed: 19390170]

Carlyon RP. Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. J Acoust Soc Am. 1996; 99:517–24. [PubMed: 8568039]

Carroll J, Zeng FG. Fundamental frequency discrimination and speech perception in noise in cochlear implant simulations. Hear Res. 2007; 231:42–53. [PubMed: 17604581]

Chang JE, Bai JY, Zeng FG. Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise. IEEE Trans Biomed Eng. 2006; 53:2598–601. [PubMed: 17152439]

Chen F, Loizou PC. Contribution of consonant landmarks to speech recognition in simulated acoustic-electric hearing. Ear Hear. 2010; 31:259–67. [PubMed: 20081538]

Ching TY, Incerti P, Hill M. Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. Ear Hear. 2004; 25:9–21. [PubMed: 14770014]

Culling JF, Darwin CJ. Perceptual separation of simultaneous vowels: within and across-formant grouping by F0. J Acoust Soc Am. 1993; 93:3454–3467. [PubMed: 8326071]

Cullington HE, Zeng FG. Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination, and talker identification. Ear Hear. 2011; 32:16–30. [PubMed: 21178567]

Dorman MF, Spahr AJ, Loizou PC, Dana CJ, Schmidt JS. Acoustic simulations of combined electric and acoustic hearing (EAS). Ear Hear. 2005; 26:371–80. [PubMed: 16079632]

Faulkner A, Rosen S, Moore BC. Residual frequency selectivity in the profoundly hearing-impaired listener. Br J Audiol. 1990; 24:381–92. [PubMed: 2279196]

Fletcher, H. Speech and Hearing in Communication Krieger. Huntington, NY: 1953.

Fournier, JE. Audiométrie vocale. Maloine, Paris: 1951.

Gantz BJ, Turner CW. Combining acoustic and electrical hearing. Laryngoscope. 2003; 113:1726–30. [PubMed: 14520097]

Gantz BJ, Turner CW. Combining acoustic and electrical speech processing: Iowa/Nucleus hybrid implant. Acta Otolaryngol. 2004; 124:344–7. [PubMed: 15224850]

Gantz BJ, Turner CW, Gfeller K. Expanding cochlear implant technology: Combined electrical and acoustical speech processing. Cochlear Implants Int. 2004; 5(Suppl 1):8–14. [PubMed: 18792215]

Gantz BJ, Turner CW, Gfeller KE. Acoustic plus electric speech processing: preliminary results of a multicenter clinical trial of the Iowa/Nucleus Hybrid implant. Audiol Neurootol. 2006; 11(Suppl 1):63–8. [PubMed: 17063013]

Gelman, A.; Carlin, J.; Stern, H.; Rubin, D. Bayesian Data Analysis. Chapman and Hall/CRC; Boca Raton, Florida: 1995.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Gfeller KE, Olszewski C, Turner C, Gantz B, Oleson J. Music perception with cochlear implants and residual hearing. Audiol Neurootol. 2006; 11(Suppl 1):12–5. [PubMed: 17063005]

Green, DM.; Birdsall, TG. Technical Memorandum No. 81 and Technical Note AFCRC-TR-57-58. University of Michigan: Electronic Defense Group; 1958. The effect of vocabulary size on articulation score.

Green DM, Dai HP. Probability of being correct with 1 of M orthogonal signals. Percept Psychophys. 1991; 49:100–1. [PubMed: 2011448]

Green, DM.; Swets, JA. Signal Detection Theory and Psychophysics. Krieger; New York: 1966.

Jaynes, ET. Probability Theory: The Logic of Science. Cambridge University Press; New York: 2003.

Jeffreys, H. The Theory of Probability. 3. Clarendon Press; Oxford: 1961.

Jordan, MI. Learning in Graphical Models. MIT Press; Cambridge, MA: 1999.

Kong YY, Braida LD. Cross-frequency Integration for Consonant and Vowel Identification in Bimodal Hearing. J Speech Lang Hear Res. 2010 in press.

Kong YY, Carlyon RP. Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. J Acoust Soc Am. 2007; 121:3717–27. [PubMed: 17552722]

Kong YY, Stickney GS, Zeng FG. Speech and melody recognition in binaurally combined acoustic and electric hearing. J Acoust Soc Am. 2005; 117:1351–61. [PubMed: 15807023]

Kryter KD. Methods for the calculation and use of the articulation index. J Acoust Soc Am. 1962; 34:1689–1697.

Kujawa SG, Liberman MC. Adding insult to injury: cochlear nerve degeneration after “temporary” noise-induced hearing loss. J Neurosci. 2009; 29:14077–85. [PubMed: 19906956]

Li N, Loizou PC. A glimpsing account for the benefit of simulated combined acoustic and electric hearing. J Acoust Soc Am. 2008; 123:2287–94. [PubMed: 18397033]

MacKay, DJC. Information theory, inference, and learning algorithms. Cambridge University Press; Cambridge, UK: 2003.

Macmillan, NA.; Creelman, CD. Detection theory: A user’s guide. 2. Erlbaum; Mahwah, NJ: 2005.

Micheyl C, Oxenham AJ. Revisiting evidence for “super-additive” benefits of combined stimulation. submitted.

Mok M, Grayden D, Dowell RC, Lawrence D. Speech perception for adults who use hearing aids in conjunction with cochlear implants in opposite ears. J Speech Lang Hear Res. 2006; 49:338–51. [PubMed: 16671848]

Moore BC. Frequency selectivity and temporal resolution in normal and hearing-impaired listeners. Br J Audiol. 1985; 19:189–201. [PubMed: 3904877]

Moore BC, Vickers DA, Glasberg BR, Baer T. Comparison of real and simulated hearing impairment in subjects with unilateral and bilateral cochlear hearing loss. Br J Audiol. 1997; 31:227–45. [PubMed: 9307819]

Müsch H, Buus S. Using statistical decision theory to predict speech intelligibility. I. Model structure. J Acoust Soc Am. 2001; 109:2896–909. [PubMed: 11425132]

Nittrouer S, Boothroyd A. Context effects in phoneme and word recognition by young children and older adults. J Acoust Soc Am. 1990; 87:2705–15. [PubMed: 2373804]

Pelli DG. Uncertainty explains many aspects of visual contrast detection and discrimination. J Opt Soc Am A. 1985; 2:1508–32. [PubMed: 4045584]

Pirenne MH. Binocular and uniocular threshold of vision. Nature. 1943; 152:698–699.

Qin MK, Oxenham AJ. Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech. J Acoust Soc Am. 2006; 119:2417–26. [PubMed: 16642854]

Ronan D, Dix AK, Shah P, Braida LD. Integration across frequency bands for consonant identification. J Acoust Soc Am. 2004; 116:1749–62. [PubMed: 15478442]

Rouder JN, Lu J. An introduction to Bayesian hierarchical models with an application in the theory of signal detection. Psychon Bull Rev. 2005; 12:573–604. [PubMed: 16447374]

Treisman M. Combining information: probability summation and probability averaging in detection and discrimination. Psychological Methods. 1998; 3:252–265.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Turner CW, Gantz BJ, Vidal C, Behrens A, Henry BA. Speech recognition in noise for cochlear implant listeners: benefits of residual acoustic hearing. J Acoust Soc Am. 2004; 115:1729–35. [PubMed: 15101651]

Tyler, RS. Frequency resolution in hearing-impaired listeners. In: Moore, BCJ., editor. Frequency Selectivity in Hearing. Academic Press; London: 1986. p. 309-371.

Uchanski RM, Braida LD. Effects of token variability on our ability to distinguish between vowels. Percept Psychophys. 1998; 60:533–43. [PubMed: 9628988]

von Ilberg C, Kiefer J, Tillein J, Pfenningdorff T, Hartmann R, Stürzebecher E, Klinke R. Electric-acoustic stimulation of the auditory system. New technology for severe hearing loss. J Otorhinolaryngol Relat Spec. 1999; 61:334–340.

Wickens, T. Elementary Signal Detection Theory. Oxford University Press; Oxford: 2001.

Appendix: Bayesian model comparison

In Bayesian model comparison, the posterior probabilities of two models, M_A and M_B, given the data, D, are compared by forming the ratio (Gelman *et al.*, 1995; Jaynes, 2003),

$$\frac{P(M_A|D)}{P(M_B|D)}=\frac{P(D|M_A)P(M_A)}{P(D|M_B)P(M_B)} \tag{A1}$$

where $P(D|M_A)$ and $P(D|M_B)$ are the conditional probabilities of the data given the model, and $P(M_A)$ and $P(M_B)$ are the model prior probabilities. When none of the models is favored *a priori* (as was the case in the current application), the prior probabilities are equal, and the ratio of posterior probabilities, $P(M_A|D)/P(M_B|D)$, equals the ratio of likelihoods, $P(D|M_A)/P(D|M_B)$. The latter ratio, known as the Bayes factor, is computed by integrating over model parameters, for each model:

$$\frac{P(D|M_A)}{P(D|M_B)}=\frac{\int_{\theta_A} P(D|\theta_A, M_A)P(\theta_A|M_A)d\theta_A}{\int_{\theta_B} P(D|\theta_B, M_B)P(\theta_B|M_B)d\theta_B} \tag{A2}$$

In the current application of this framework, the data were numbers of correct responses for lowpass, vocoded, and combined (lowpass + vocoded) stimuli. These numbers of correct responses are denoted as $n_L^{i,j}$, $n_V^{i,j}$, and $n_C^{i,j}$ (where the superscripts, i and j, index the listener and the condition), and are assumed to be drawn from binomial distributions with parameters, n, $p_L^{i,j}$, $p_V^{i,j}$, and $p_C^{i,j}$. For clarity, and using probability-theory notation, we have:

$$n_L^{i,j}|p_L^{i,j}, n \sim binomial\left(p_L^{i,j}, n\right), \tag{A3}$$

$$n_V^{i,j}|p_V^{i,j}, n \sim binomial\left(p_V^{i,j}, n\right), \tag{A4}$$

and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$n_C^{i,j} | p_C^{i,j}, n \sim binomial\left(p_C^{i,j}, n\right). \tag{A5}$$

The shared parameter, n, corresponds the number of trials per condition per listener; this number if fixed and known (n= 20). The parameters, $p_L^{i,j}$, $p_V^{i,j}$, and $p_C^{i,j}$, are the underlying (or "true") probabilities of a correct response. These probabilities are "latent" variables; they are not observed. Although their most-likely (or maximum-*a-posteriori*) values can be inferred based on the measured numbers of correct responses, for the purpose of evaluating the right-hand side of Eq. A2, these variables are just "nuisance parameters," which must be integrated over to compute the conditional probabilities of interest.

The different models described in section 2.4 define specific relationships between $p_L^{i,j}$, $p_V^{i,j}$, and $p_C^{i,j}$. Specifically, for the probability-summation model,

$$p_C^{i,j} = 1 - \left(1 - p_L^{i,j}\right)\left(1 - p_V^{i,j}\right). \tag{A6}$$

For the independent-noises model,

$$p_C^{i,j} = f_m^{-1}\left(\sqrt{f_m^2\left(p_L^{i,j}\right) + f_m^2\left(p_V^{i,j}\right)}\right). \tag{A7}$$

where $f_m(.)$ denotes the transformation from d' to PC for the mAFC task, as defined by the integral Eq. 3, and $f_m^{-1}(\cdot)$ denotes the inverse of this transformation. Finally, for the late-noise model,

$$p_C^{i,j} = f_m^{-1}\left(f_m\left(p_L^{i,j}\right) + f_m\left(p_V^{i,j}\right)\right). \tag{A8}$$

In general terms, we can write,

$$p_C^{i,j} = g_k\left(p_L^{i,j}, p_V^{i,j}\right). \tag{A9}$$

where the subscript, k, refers to the considered model (1: probability-summation model; 2: independent-noises model; 3: late-noise model).

Together with Eq. A9, and with the additional assumption that $p_L^{i,j}$ and $p_V^{i,j}$ are marginally independent, the conditional-dependence statements A3–A5 define a directed acyclic graph (Bishop, 2006; Jordan, 1999). By examining the structure of this graph, it can be determined that, for the probability-summation model,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$P\left(N_L^{i,j}, N_V^{i,j}, N_C^{i,j}|M_k\right) = \int_0^1 B_n\left(n_L^{i,j}, x_L\right) P(x_L) \int_0^1 B_n\left(n_V^{i,j}, x_V\right) P(x_V) B_n\left(n_C^{i,j}, f_k(x_L, x_V)\right) dx_L dx_V, \quad k=1, \tag{A10}$$

where $N_L^{i,j}$, $N_V^{i,j}$, and $N_C^{i,j}$ denote the random variables, of which $n_L^{i,j}$, $n_V^{i,j}$, and $n_C^{i,j}$ are observed realizations; x_L and x_V are "dummy" integration variables, which correspond to $p_L^{i,j}$ and $p_V^{i,j}$; and $B_n(q, p)$ is the probability distribution function with parameters n (the number of trials) and p (the probability of success) evaluated at q (the number of successes),

$$B_n(q, p) = \frac{n!}{q!(n-q!)} p^q (1-p)^{n-q}. \tag{A11}$$

Uniform prior distributions were placed on parameters $p_L^{i,j}$ and $p_V^{i,j}$. Consequently, the terms $P(x_L)$ and $P(x_V)$ in Eq. A11 were both equal to one, and,

$$P\left(N_L^{i,j}, N_V^{i,j}, N_C^{i,j}|M_k\right) = \int_0^1 B_n\left(n_L^{i,j}, x_L\right) \int_0^1 B_n\left(n_V^{i,j}, x_V\right) B_n\left(n_C^{i,j}, f_k(x_L, x_V)\right) dx_L dx_V, \; k=1. \tag{A12}$$

The independent-noises model and the late-noise model have one more parameter than the probability-summation model. This parameter, m, enters in the transformation between d' and PC for the mAFC task (Eq. 3). It corresponds to the number of independent templates, against which the listener is comparing incoming speech signals, and can be thought of as the size of the listener's "active vocabulary" (Green & Birdsall, 1958; Müsch & Buus, 2001). Since m is not known *a priori*, this extra parameter must also be "integrated out" in the calculation of the likelihoods of the independent-noises and late-noise models. Accordingly, for these models, the likelihood was computed as follows.

$$P\left(N_L^{i,j}, N_V^{i,j}, N_C^{i,j}|M_k\right) = \sum_0^\infty P(m) \int_0^1 B_n\left(n_L^{i,j}, x_L\right) \int_0^1 B_n\left(n_V^{i,j}, x_V\right) B_n\left(n_C^{i,j}, f_k(x_L, x_V)\right) dx_L dx_V dm, \quad k=\{2, 3\}. \tag{A13}$$

The infinite upper limit in the sum on the right-hand side of Eq. A13 reflects the fact that m can theoretically be any non-negative integer. However, the size of listeners' active vocabulary is most certainly finite. For practical purposes, uncertainty concerning this parameter was modeled using a uniform-discrete prior probability distribution on m, such that values of m equally spaced on an octave scale between 125 and 125×2^{13} were regarded as equally likely *a priori*.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The integrals in Eqs. A12 and A13 were replaced by sums, and evaluated numerically. The Bayes factor were then obtained simply by taking ratios of the conditional probabilities, $P\left(N_L^{i,j}, N_V^{i,j}, N_C^{i,j}|M_k\right)$, corresponding to two different models (e.g., the probability-summation model and the independent-noises model). This was done for all pairs of models. Importantly, these Bayes factors were computed separately for each listener and condition. The mean Bayes factors reported in the Results section were computed as the geometric mean, across all listeners, of these individual Bayes factors.

In addition to "individual" Bayes factors, "group-level Bayes factors" were computed as

$$\frac{P(D|M_A)}{P(D|M_B)} = \frac{\prod_{j=1}^{n_{subjs}} P\left(N_L^{i,j}, N_V^{i,j}, N_C^{i,j}|M_A\right)}{\prod_{j=1}^{n_{subjs}} P\left(N_L^{i,j}, N_V^{i,j}, N_C^{i,j}|M_B\right)}, \tag{A14}$$

where n_{subjs} is the number of subjects, and M_A and M_B refer to the two models being compared.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Highlights

- Combining lowpass-filtered and vocoded signals improves speech-recognition performance
- This effect is observed under a wide variety of stimulation conditions
- It is larger than predicted by a model combining independent identification decisions
- Gaussian-SDT models can explain the effect without involving cross-modal interactions

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

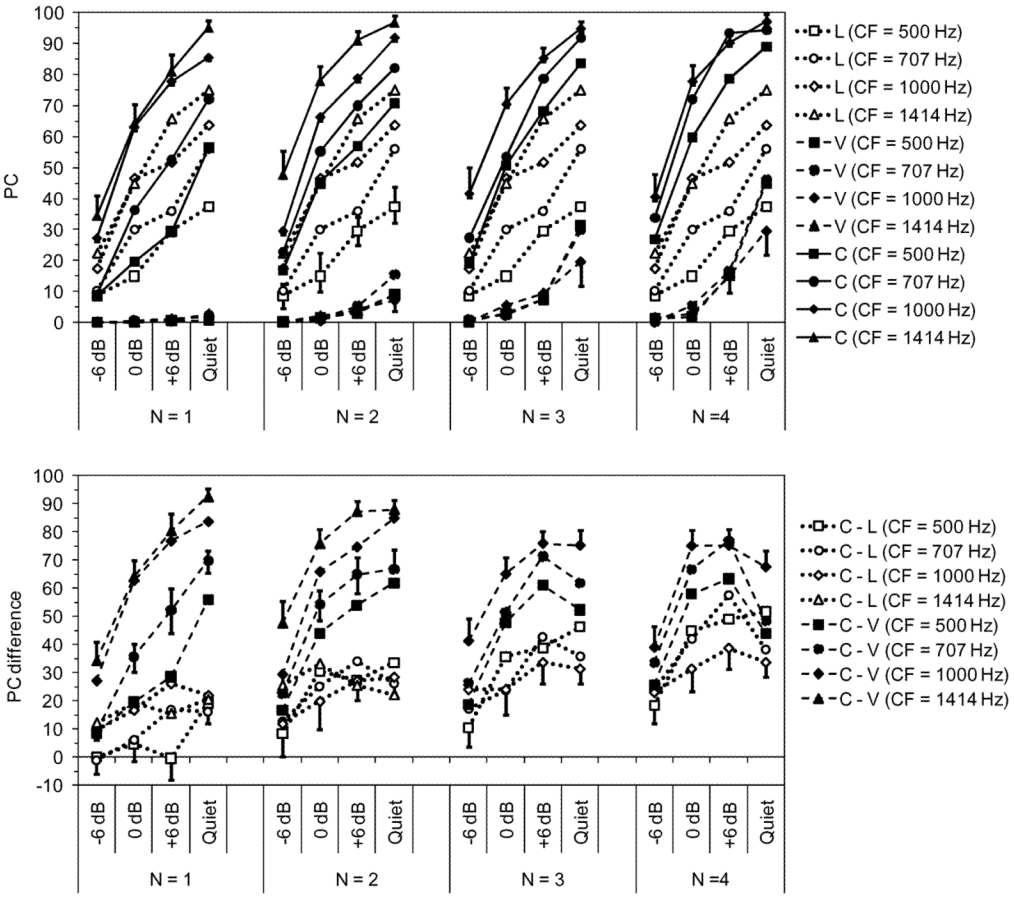

Figure 1.

PC in L, V, and C conditions, and combined-stimulation advantage, as a function of N and SNR, with CF as the parameter. The different CF conditions are illustrated by different symbols, as indicated in the key. The different N and SNR conditions are listed underneath the abscissa. Note that N refers to conditions that involved vocoded stimuli. The data obtained in conditions involving only lowpass-filtered stimuli (empty symbols in the upper panel) were duplicated across N conditions in the upper panel, to facilitate comparisons. The PC differences, which are shown in the lower panel, were computed by subtracting the PCs measured in the L or V conditions from the PCs measured in corresponding C conditions. The error bars show the upper and/or lower bounds of the 95%-confidence intervals (bootstrap) of the mean PC across the 24 listeners for a subset of conditions. These conditions were selected so as to minimize overlap and avoid clutter. They include conditions in which the mean PCs were the lowest or the highest in each N condition; the 500-Hz CF for the N = 2 condition (L) in the upper panel; and the 707-Hz CF for N = 1 and N = 2 condition (C – V) in the lower panel. Note that some error bars for this subset of conditions are not visible, due to their small size relative to the symbols.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

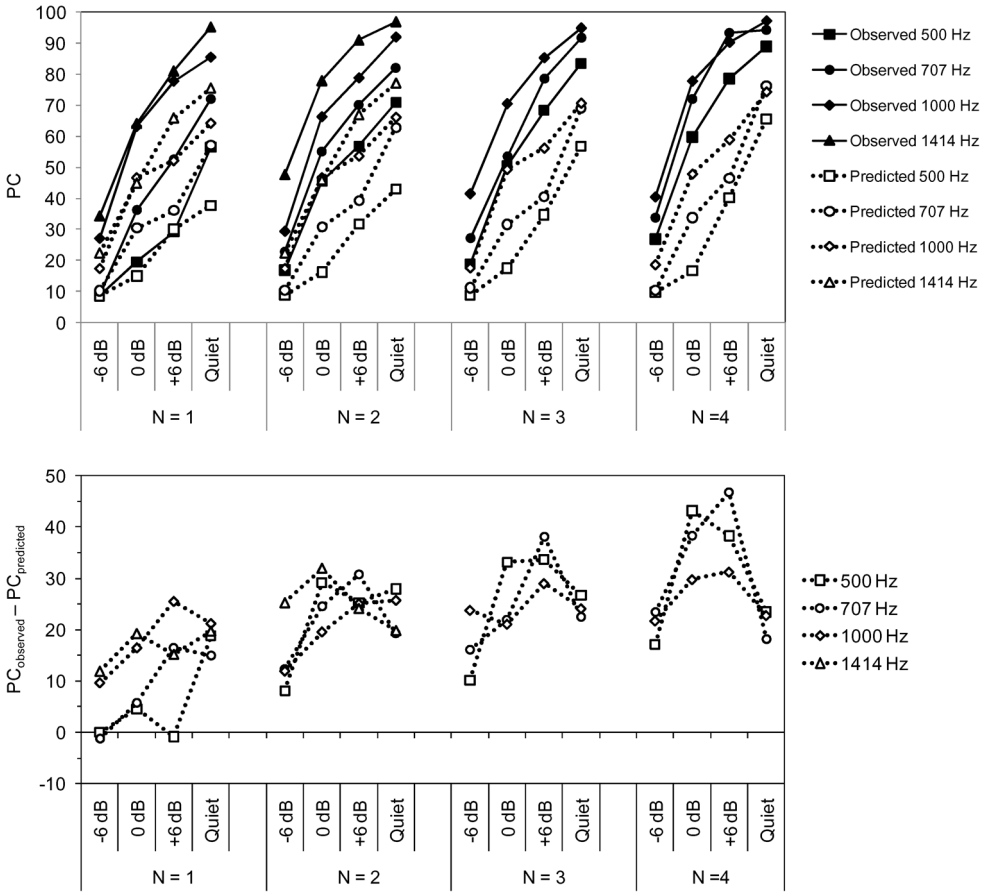

Figure 2.

Comparison between the PCs measured for C stimuli in human listeners, and the PCs predicted by the probability-summation model. Upper panel: Mean PCs of the human listeners for C stimuli (filled symbols, replotted from Fig. 1), and model predictions (empty symbols). Lower panel: Differences between observed and predicted PCs. These differences were obtained by subtracting the predicted PCs from the observed PCs, so that positive values indicate conditions in which listeners' performance was better than predicted by the model, and negative values indicate the opposite effect.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

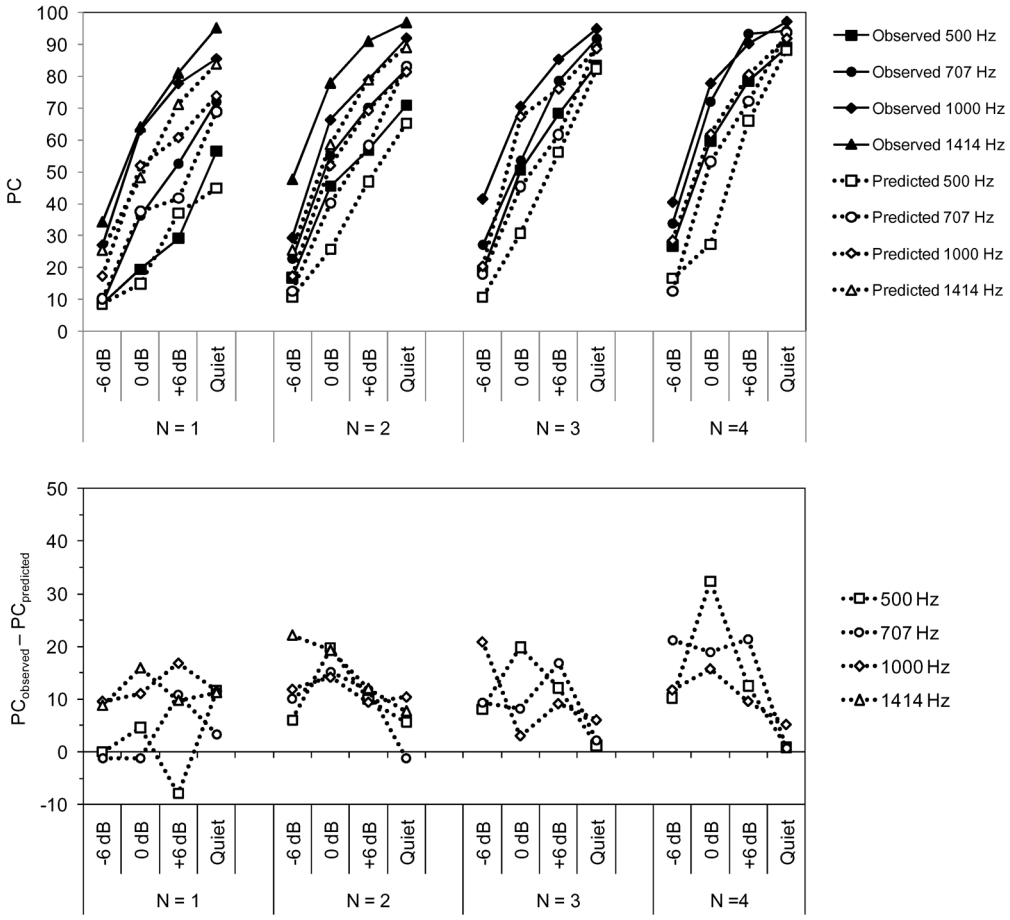

Figure 3.

Comparison between human listeners' performance and the predictions of the independent-noises model, with the value of the parameter, *m*, set to 8000. The format of this figure is the same as that of Fig. 2.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

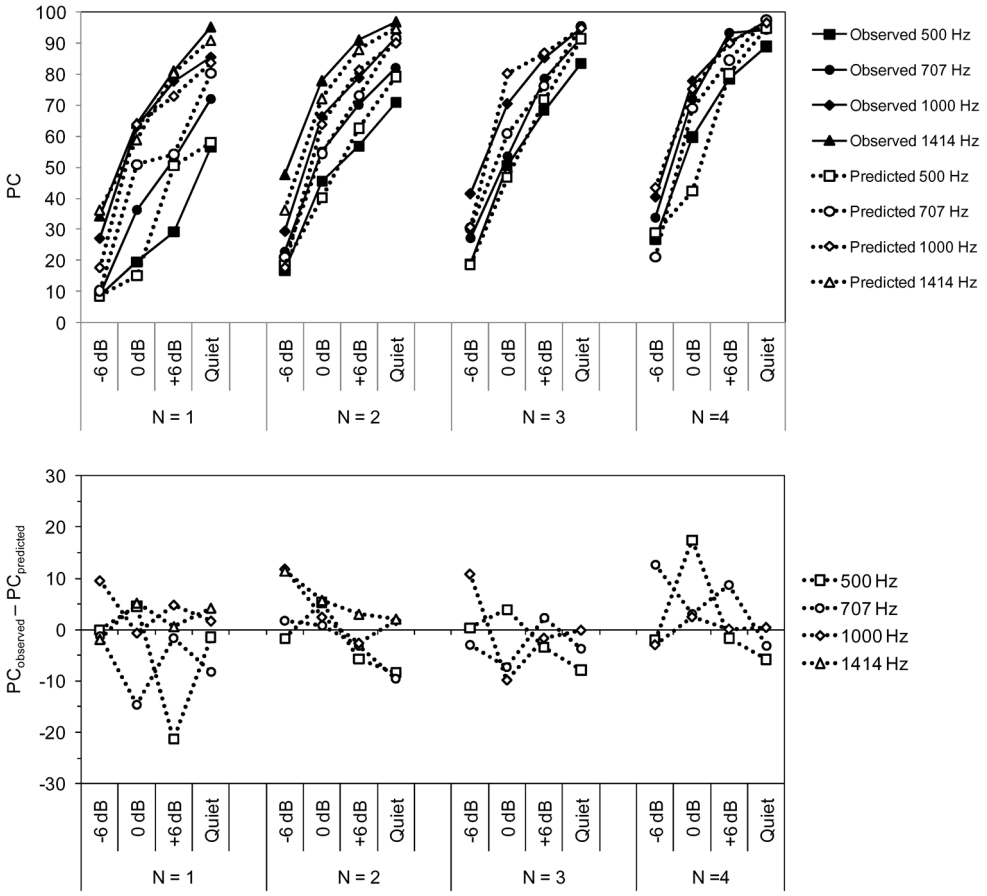

Figure 4.

Comparison between human listeners' performance and the predictions of the independent-noises model, with the value of the parameter, m, adjusted to minimize the mean-squared error between data and predictions. The format of this figure is the same as that of Fig. 2.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

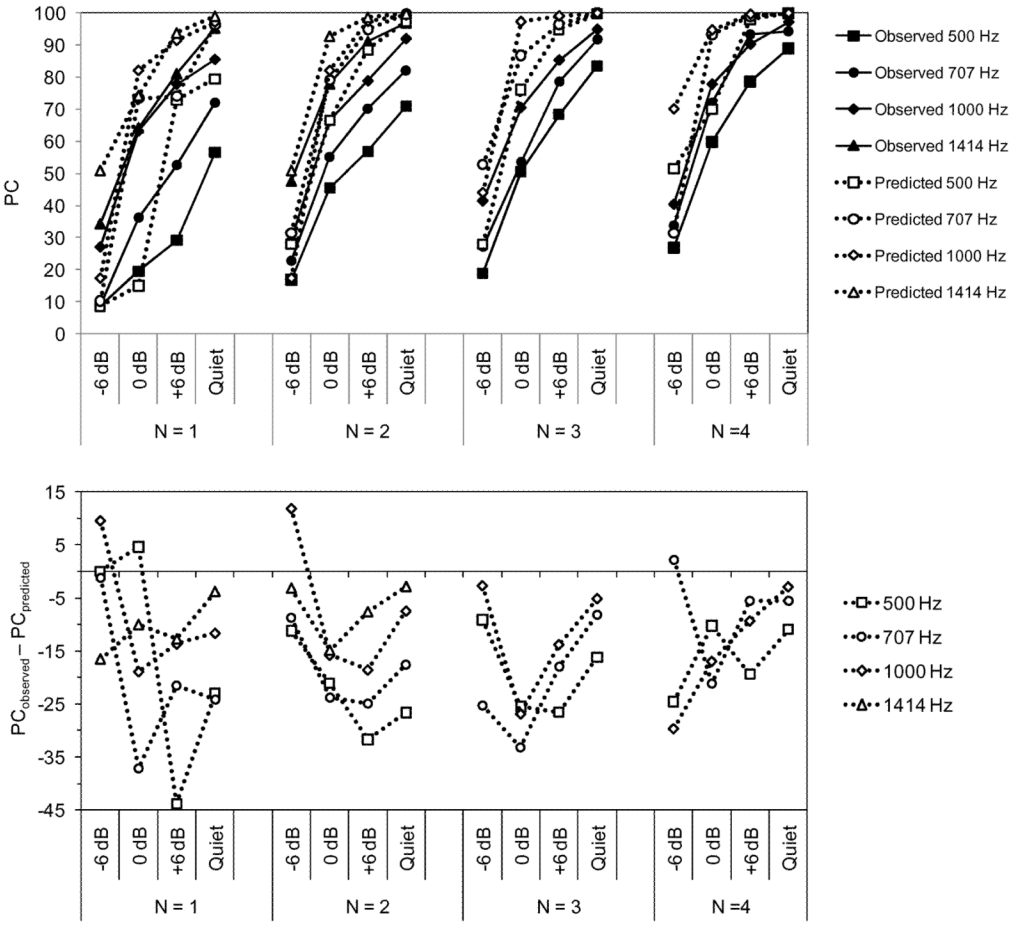

Figure 5.

Comparison between human listeners' performance and the predictions of the late-noise model with the value of the parameter, m, set to 8000. The format of this figure is the same as that of Fig. 2.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

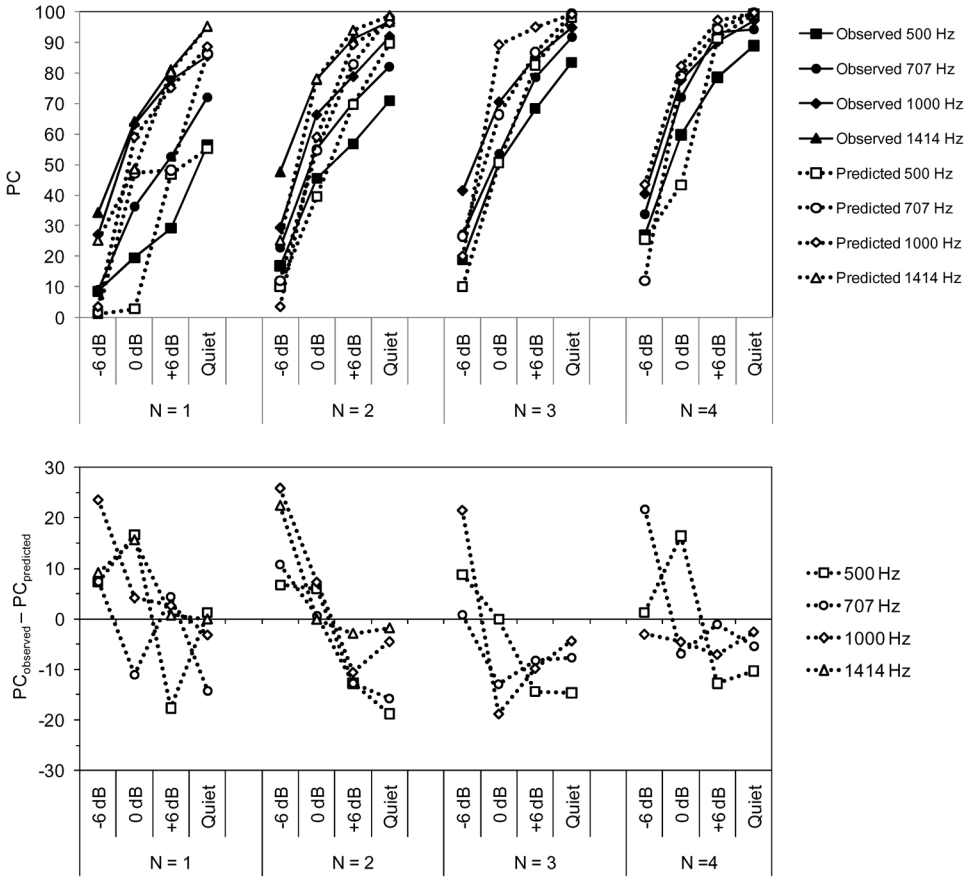

Figure 6.

Comparison between human listeners' performance and the predictions of the late-noise model with the value of the parameter, m, adjusted to minimize the mean-squared error between data and predictions. The format of this figure is the same as that of Fig. 2.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

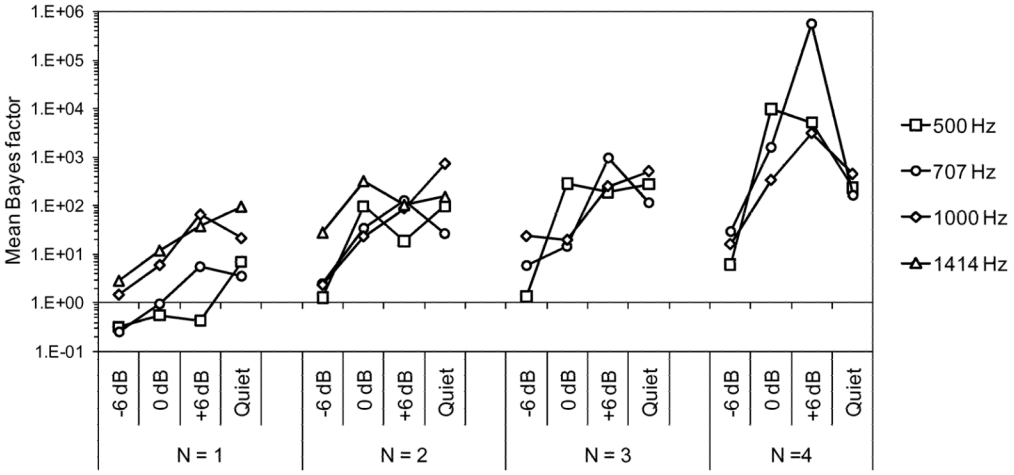

Figure 7.

Geometric-mean Bayes factors for the comparison between the independent-noises and probability-summation models. Values higher than 1 (indicated by the horizontal line) indicate that, on average across all listeners, for the considered combination of N, SNR, and CF, the data measured in the experiment were more likely to have been generated by the independent-noises model than by the probability-summation model; values lower than 1 indicate the opposite. Importantly, these Bayes factors refer to *individual* data; the Bayes factors that were computed on the group data (assuming statistical independence across listeners) were orders of magnitude higher.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table I

Low- and high-frequency limits of the vocoder synthesis bands for the different CF and N conditions.

Condition	Band number	Synthesis-band limits (Hz)
CF = 500 Hz		
N = 1	1	575 – 3925
N = 2	1	575 – 1339
	2	1489 – 3925
N = 3	1	575 – 925
	2	1075 – 1925
	3	2075 – 3925
N = 4	1	575 – 766
	2	916 – 1339
	3	1489 – 2303
	4	2453 – 3925
CF = 707 Hz		
N = 1	1	782 – 3925
N = 2	1	782 – 1607
	2	1757 – 3925
N = 3	1	782 – 1185
	2	1335 – 2170
	3	2320 – 3925
N = 4	1	782 – 1015
	2	1165 – 1606
	3	1756 – 2517
	4	2667 – 3925
CF = 1000 Hz		
N = 1	1	1075 – 3925
N = 2	1	1075 – 1925
	2	2075 – 3925
N = 3	1	1075 – 1512
	2	1662 – 2444
	3	2594 – 3925
N = 4	1	1075 – 1339
	2	1489 – 1925
	3	2075 – 2753
	4	2903 – 3925
CF = 1414 Hz		
N = 1	1	1489 – 3925
N = 2	1	1489 – 2303
	2	2453 – 3925