MIT Open Access Articles

Masking Release for Igbo and English

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Ebem, D. U., J. G. Desloge, C. M. Reed, L. D. Braida, and J. O. Oguru. "Masking Release for Igbo and English." African Journal of Computing & ICT 6(3) (September 2013): 119-126.

As Published: http://www.ajocict.net/vol_6_no_3_september_2013.html

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: http://hdl.handle.net/1721.1/99894

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike

Massachusetts Institute of Technology

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH Public Access

Author Manuscript

Afr J Comput Ict. Author manuscript; available in PMC 2014 October 21.

Published in final edited form as:

Afr J Comput Ict. 2013 September ; 6(3): 119–126.

Masking Release for Igbo and English

Deborah U. Ebem, PhD[1], **Joseph G. Desloge, PhD**[2], **Charlotte M. Reed, PhD**[2], **Louis D. Braida, PhD**[2], and **Joy O. Uguru, PhD**[3]

Deborah U. Ebem: deborah.ebem@unn.edu.ng; Charlotte M. Reed: cmreed@mit.edu; Louis D. Braida: ldbraida@mit.edu; Joy O. Uguru: joy.uguru@unn.edu.ng

[1]Department of Computer Science, University of Nigeria, Nsukka.

[2]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts U.S.A.

[3]Department of Linguistics, Igbo and Other Nigerian Languages, University of Nigeria, Nsukka. Nigeria.

Abstract

In this research, we explored the effect of noise interruption rate on speech intelligibility. Specifically, we used the Hearing In Noise Test (HINT) procedure with the original HINT stimuli (English) and Igbo stimuli to assess speech reception ability in interrupted noise. For a given noise level, the HINT test provides an estimate of the signal-to-noise ratio (SNR) required for 50%-correct speech intelligibility. The SNR for 50%-correct intelligibility changes depending upon the interruption rate of the noise. This phenomenon (called Masking Release) has been studied extensively in English but not for Igbo – which is an African tonal language spoken predominantly in South Eastern Nigeria. This experiment explored and compared the phenomenon of Masking Release for (i) native English speakers listening to English, (ii) native Igbo speakers listening to English, and (iii) native Igbo speakers listening to Igbo. Since Igbo is a tonal language and English is a non-tonal language, this allowed us to compare Masking Release patterns on native speakers of tonal and non-tonal languages.

Our results for native English speakers listening to English HINT show that the SNR and the masking release are orderly and consistent with other English HINT data for English speakers. Our result for Igbo speakers listening to English HINT sentences show that there is greater variability in results across the different Igbo listeners than across the English listeners. This result likely reflects different levels of ability in the English language across the Igbo listeners. The masking release values in dB are less than for English listeners.

Our results for Igbo speakers listening to Igbo show that in general, the SNRs for Igbo sentences are lower than for English/English and Igbo/English. This means that the Igbo listeners could understand 50% of the Igbo sentences at SNRs less than those required for English sentences by either native or non-native listeners. This result can be explained by the fact that the perception of Igbo utterances by Igbo subjects may have been aided by the prediction of tonal and vowel harmony features existent in the Igbo language. In agreement with other studies, our results also show that in a noisy environment listeners are able to perceive their native language better than a second language. The ability of native language speakers to perceive their language better than a second language in a noisy environment may be attributed to the fact that:

a. Native speakers are more familiar with the sounds of their language than second language speakers.

b. One of the features of language is that it is predictable hence even in noise a native speaker may be able to predict a succeeding word that is scarcely audible. These contextual effects are facilitated by familiarity.

Keywords

masking release; intilligibility; interruption rate; hearing in noise test

1. Introduction

In this experiment, we are exploring the effect of interruption rate on intelligibility. Specifically, we are using the Hearing In Noise Test (HINT)[[1]–[3]] procedure with either the original HINT stimuli (English) or our own stimuli (Igbo) to assess speech reception ability in interrupted noise. For a given noise level, the HINT test provides an estimate of the signal-to-noise ratio (SNR) required for 50%-correct speech intelligibility. This signal level (for 50%-correct intelligibility) changes depending upon the interruption rate of the noise. This phenomenon (called Masking Release) has been studied extensively in English [[4]–[12]] but not for Igbo – which is an African tonal language spoken predominantly in South Eastern Nigeria. This experiment will explore and compare the phenomenon of Masking Release for (i) native English speakers listening to English, (ii) native Igbo speakers listening to English, and (iii) native Igbo speakers listening to Igbo. Since Igbo is a tonal language and English is a non-tonal language, this will allow us to compare Masking Release patterns on native speakers of tonal and non-tonal languages.

Three main experiments were carried out using the original HINT stimuli (English) and our own stimuli (Igbo). In the first experiment (Experiment 1) native English-speaking subjects listened to the original HINT stimuli at the Sensory Communication Laboratory at Massachusetts Institute of Technology. In the second experiment (Experiment 2) native Igbo-speaking subjects having a good understanding of English listened to the same HINT stimuli. In the third experiment (Experiment 3) the same Igbo-speaking subjects listened to the original (undegraded) Igbo speech stimuli [13]. The first experiment was carried out in the United States, and the other two experiments in Nigeria. The Nigerian subjects were native Igbo speakers proficient in English.

The results of these three main experiments will allow us to compare Masking Release patterns on native speakers of tonal and non-tonal languages, since Igbo is a tonal language and English is a non-tonal language. In addition, the experiment allows us to compare the performance of native with non-native English speakers on the English HINT test.

The remainder of this paper is organized as follows: Section 2 presents the methodology. In Section 3 the results of the experiments are presented. Section 4 discusses the results of the experiments. The paper is concluded in Section 5.

2. Methodology

The same basic test methodology was employed in Experiments 1, 2, and 3 to measure SNR for 50%-correct sentence intelligibility as a function of the interruption rate of the noise. For these adaptive measurements of SNR, the noise level was always 65 dB SPL and the level of the speech stimuli was varied to measure the 50%-correct level of performance. The noises used for the English and the Igbo sentences had speech-shaped spectra which were matched to the long-term spectrum of the recorded English and Igbo sentences, respectively. SNRs for 50%-correct sentence reception were measured for four values of temporally interrupted noise: 4, 8, 10, and 20 Hz and for continuous (i.e., uninterrupted) noise. The interrupted-noise was created using square-wave modulation (at 4, 8, 10, and 20 Hz) at a modulation depth that yielded a 30-dB SPL noise level in the troughs. The level of the interrupted noises was adjusted to have the same overall level of 65 dB SPL as the steady-state, uninterrupted noise. All sentences were presented monaurally under HD 202 Sennheiser headphones in a sound-treated room. The speech was played directly from a Dell laptop computer.

The subject’s task in these experiments was to listen to sentences as they were presented and to repeat aloud whatever had been heard or understood. Prior to conducting the main experiments, a verification procedure was employed to determine if each subject’s responses could be scored consistently by two independent experimenters. The verification procedure involved two lists of English HINT sentences spoken by a male talker (that were not used again in the main Experiments 1 and 2) and two lists of Igbo sentences (from a female talker not used in the main Experiment 3). For the verification procedure, the sentences were presented in a noise of 65-dB SPL at a constant signal-to-noise ratio (SNR) of 0 dB. Following each sentence presentation, the subject was asked to repeat what he or she had heard and two experimenters recorded the responses independently. If both experimenters were in agreement in scoring the subject’s responses 90% or more of the time, then the subject continued with the experiment. This criterion was met for each of the subjects in each of the three experiments.

In the three main experiments, an adaptive procedure was used to determine the SNR at which 50% of the sentences could be repeated correctly. Each individual run was conducted with a list of 10 sentences. The first sentence of a given list was repeated on consecutive trials with increasing level (in 8 dB steps) until the subject was able to repeat it correctly. The nine remaining sentences in the list were presented once each, with the presentation level increased after an incorrect response or decreased following a correct response. The step size for changing level was 4 dB until the first “reversal” in level, and then was 2 dB for the rest of the run. The SNR for 50%-correct sentence reception was obtained by averaging the SNR for the final six sentences in the list. During HINT testing, an experimenter was present in the sound-treated room with the subject. The subject was asked to provide a word-for-word oral response each stimulus. The experimenter scored each response as either ‘correct’ if the subject identified all words correctly (with minor exceptions such as a/the and is/was) or ‘incorrect’ otherwise.

A comparison by the experimenter of the subject’s response to the text version of the sentence was used to judge accuracy. The text of the sentences listed specific, small variations in the sentences that were to be allowed as correct responses based upon responses repeatedly made by subjects during the previous testing. The computer recorded sentence presentation levels, whether the subject's response was correct (as determined by the experimenter), as well as the identity of the specific sentence on each trial.

3. Experiments and Results

3.1 Experiment 1: Native English Subjects with English HINT Sentences

A total of six native American English speaking males (2) and females (4) were paid for their participation. The experiment was carried out at the Sensory Communication Laboratory of Massachusetts Institute of Technology Cambridge. All subjects were screened for normal hearing (15 dB Hearing Level (HL) from 0.25 to 8 kHz).

3.1.i. English HINT test—This testing employed the HINT sentences which consist of 26 phonetically balanced 10-sentence lists. Two of the lists were used for the response verification procedure discussed above, four lists were used for practice on the experimental conditions prior to testing, and twenty of the lists were used in the main experiment. For each subject, the experiment involved four independent SNR measurements for each of the five different noise conditions. The order of the five conditions was randomized for each replication and one list of sentences was presented for each condition under each test order. Within a given list of sentences, the order in which sentences was presented was randomized for each subject. Each list of sentences was used only once during the main experiment, since listeners can remember what they have heard from previous presentations.

3.1.ii. Results—Figure 1 shows results for English listeners with English HINT sentences. There is low variability among the six subjects. The mean SNR (in upper plot) is similar for 4 and 8 Hz (roughly -11 dB) and then increases at rates of 32 and 64 Hz and for continuous noise. The masking release, defined as the SNR in continuous noise minus the SNR in interrupted noise (lower plot), is about 10 dB for the 4 and 8-Hz conditions, 5 dB for 32-Hz, and 1 dB for 64 Hz. These data are orderly and are consistent with other English HINT data for English speakers.

3.2. Experiment 2: Native Igbo Subjects with English HINT Sentences

A total of seven native Igbo speaking males (4) and females (3) who are proficient in English participated in the experiment using the same HINT corpus and the same procedure was followed as that done by the native English speakers at MIT. The experiment was conducted in Enugu, Nigeria.

The verification procedure was first performed to ensure that the subjects' responses could be scored consistently across two experimenters. The details of the experiment followed those described above for Native English subjects.

3.2.i. Results—Figure 2 shows results for Igbo listeners with English HINT sentences. Greater variability was observed in the SNR measures across the different Igbo listeners than across the English listeners. This finding likely reflects different levels of ability in the English language across the Igbo listeners. Subjects 4, 6, and 7 performed similarly to each other and have the lowest SNRs (i.e., best performance) in the upper plot. The other four subjects generally require higher SNR. Interestingly, the subjects are most alike on the continuous noise condition, and show greatest variation for the 4 and 8 Hz interruption rates. The masking release values in dB (lower plot) are less than for English listeners; that is, the interruptions were not as beneficial in aiding speech reception for non-native listeners as they were for native listeners of English. Two of the Igbo subjects (Subjects 6 and 7) who are highly proficient in English, however, did have almost as much masking release as the English listeners.

3.3 Experiment 3: Native Igbo Subjects with Igbo Sentences

The same seven native Igbo speaking males and females who are proficient in English and participated using the HINT corpus were used in this experiment using our Igbo speech corpus [13]. The experiment was done in Enugu, Nigeria. The same general procedure and apparatus was used as described above for the HINT experiment. The Igbo speech corpus is described below.

3.3.i. The Igbo Speech Database—The Igbo speech corpus was high quality 48-kHz sampling, 16–bit resolution recording. It was recorded in a low –reverberation environment and includes sentences produced by four talkers (two males and two females). The sentences were arranged into 10-sentence lists for use in the experiment. The verification procedure used two sentences from a female list and the practice runs on the experimental noise conditions used another two lists from a female talker. The main experiment used 20 lists from the two male talkers (10 10-sentence lists per talker). The duration of the sentences was between 4 and 6 seconds, as used in the International Telecommunication Union (ITU-T) Perceptual Evaluation of Speech Quality (PESQ) and Perceptual Objective Listening Quality Assessment (POLQA) standardization efforts [14],15].

3.3.ii. Results—Figure 3 shows results for Igbo listeners with Igbo sentences. In general, the SNRs for Igbo sentences are lower than for English/English and Igbo/English. For the continuous noise, for example, the mean Igbo/Igbo SNR averaged roughly −7 dB compared to roughly −2 dB for English/English testing. That means that the Igbo listeners could understand 50% of the Igbo sentences at lower SNRs than required for English either by native or non-native listeners. The masking-release results for Igbo/Igbo, however, are similar to English/English.

3.4 Mean SNR and Masking Release results for the three groups: English/English, Igbo/English, and Igbo/Igbo

Figure 4 below shows mean SNR and masking release results for the three groups: English/English, Igbo/English, and Igbo/Igbo.

This allows us to see comparisons among the groups. In the top plot, SNR for Igbo/Igbo is lower (better) than for English/English. The non-native condition requires higher SNR at all interruption rates, as would be expected. In the lower plot, we can see that the masking release is smallest for Igbo/English and fairly similar for English/English and Igbo/Igbo.

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

One small difference is that for Igbo there is more masking release at 8 Hz than at 4 Hz, whereas there is similar masking release at 4 and 8 Hz for English/English.

A comparison of Igbo/English and Igbo/Igbo results shows that in noisy environments, listeners are able to perceive their native language better than a second language.

4. Discussion

A number of factors may have led to the differences in results between English and Igbo. One factor is differences in the test materials used for the two different languages. The English HINT sentences [1] were developed with a number of strict criteria (which have also been applied to other languages—see [2] [3]). The lists of sentences are constructed to be phonetically balanced and to be of equal difficulty. Furthermore, these sentences have been shown to have acceptable within-list response variability and inter-list reliability. Such test development strategies were not employed in constructing the Igbo sentence materials. Although these sentences represent a sample of 200 utterances from 2 different male talkers, we have not verified that the 10-sentence lists that we constructed are phonetically balanced or are equally difficult. Thus, it is possible that one factor contributing to the lower SNRs observed for Igbo/Igbo than for English/English arises from differences in the test-material construction. Although the differences in test-material construction may have contributed to the lower SNRs observed for Igbo/Igbo, these differences may have less of an effect on the relative measures of Masking Release. And, in fact, the Masking Release for Igbo/Igbo and English/English are quite comparable.

Another possible reason for the different SNRs measured for English/English and Igbo/Igbo is that perception is language specific [16]. According to [16] the phonotactics (constraints on the way phonemes co-occur), word boundary pattern or even vowel harmony can be utilized by listeners in predicting utterances, thus aiding their perception.

In this experiment, the perception of the Igbo utterances by the Igbo subjects may have been aided by the prediction of tonal and vowel harmony features existent in the Igbo language. Let us consider the possibility of tone first. In tone languages, every syllable in an utterance has a definite pitch of its own. This is contrary to what obtains in intonation languages, to which English belongs, where an utterance generally has a significant pitch located in the nuclear syllable. This therefore means that the nuclear syllable, being the most prominent syllable in the utterance, has the highest chance to be perceived relative to the other syllables. The acoustic correlates of intonation (which are identifiable with the nuclear syllable) include fundamental frequency, long vowel, loudness, long duration, stress and prominence [17]; this indicates that fundamental frequency and duration of vowels are consistent acoustic correlates of intonation. Hence in a noisy environment, a listener may only perceive this nuclear syllable and there may be little or no meaning if the other syllables of the same utterance were not perceived. Contrarily, the syllables of an utterance in a tone language like Igbo have equal prominence and thus they have equal chances of perception in any environment. If one is heard, the likelihood that others will also be heard is high and this aids meaning and communication in a noisy environment.

These arguments can be exemplified below:

Igbo: Ó tòrò ézígbó ógólógó. He is very tall.

3(sg) tall very tall

English: he is very `tall.

In the examples above, all the syllables of the Igbo utterance have tones indicated on them while only the nuclear syllable has the intonation mark indicated on it in the English utterance.

Additionally, the prediction of vowel harmony may have facilitated the perception on the Igbo utterances. Igbo has two sets of vowels as far as vowel harmony is concerned. These are:

1. +ATR vowels (e, i, o, u) /e,i,o,u/
2. –ATR vowels (a,[, q,x,) /a, ✋,❷,▯/

The Igbo utterance in the example above, has +ATR vowels. Hence one of the test utterances used – Uche na-eche oke uche (see third utterance of first male speaker) has +ATR vowels while another test utterance, nka ka anyi na-arq gi (see fourth utterance of first male speaker) is made up of –ATR vowels. In these utterances, the listener, on hearing the first syllables (if they are audible) could infer what the succeeding phonemes are. Even in test utterances containing words with – ATR vowels as well as those bearing +ATR vowels, this method of prediction can also be applied. Furthermore, prediction may have been enhanced because some of the test utterances are disyllabic words while the short speech fragments are predominantly made up of monosyllabic and disyllabic words. In Igbo, it is not possible to have words like Ucha or Uchq. Hence the listener can easily deduce what the second vowel of such a disyllabic noun could be even when the speech is not very clear. The vowel ‘u’ for instance, can co-occur with only four possibilities - /o i e u/ . This therefore makes for easier speech recognition in noisy environment.

5. Conclusions

In this paper we investigated the effect of interruption rate on speech intelligibility. Specifically, we used the Hearing In Noise Test (HINT) procedure with the original HINT stimuli (English) and Igbo stimuli. For a given noise level, the HINT test provides an estimate of the signal level required for 50%-correct speech intelligibility. This signal level (for 50%-correct intelligibility) changes depending upon the interruption rate of the noise. This phenomenon (called Masking Release) has been studied extensively in English but not for Igbo-an African tonal language spoken in predominantly in South Eastern Nigeria. This experiment explored and compared the phenomenon of Masking Release for (i) native English speakers listening to English, (ii) native Igbo speakers listening to English, and (iii) native Igbo speakers listening to Igbo. Since Igbo is a tonal language and English is a non-tonal language, this allowed us to compare Masking Release patterns on native speakers of tonal and non-tonal languages.

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

Our results for native English speakers listening to English HINT show that the SNR and the masking release are orderly and consistent with other English HINT data for English speakers. Our results for Igbo speakers listening to English HINT show that there is greater variability in results across the different Igbo listeners than across the English listeners. This could reflect different levels of ability in the English language across the Igbo listeners. The average masking release values in dB are less than for English listeners.

Our results for Igbo speakers listening to Igbo show that in general, the SNRs for Igbo sentences are lower than for English/English and Igbo/English. This means that the Igbo listeners could understand 50% of the Igbo sentences in much lower SNRs than are required for English either by native or non-native listeners. The size of the masking release for Igbo/Igbo, however, was similar to that for English/English. The perception of Igbo utterances by Igbo subjects may have been aided by the prediction of tonal and vowel harmony features existent in the Igbo language. Our results also show that in a noisy environment listeners are able to perceive their native language better than a second language. The ability for native language speakers to perceive their language better than a second language in a noisy environment may be attributed to the fact that:

a. Native speakers are more familiar with the sounds of their language than second language speakers.

b. One of the features of language is that it is predictable hence even in noise a native speaker may be able to predict a succeeding word that is scarcely audible. This is facilitated more by familiarity.

This multi-cultural voice-based experiment supports the development and design of voice systems for English-speakers and non-native English speakers using the Igbo tonal language as our test-bed. It is expected that our findings and results will have strong implications for designing voice-based systems for online communication, the design of hearing aids, and it will also determine the contributions of various properties of background noise to the speech recognition difficulties experienced by various categories of listeners when using voice-based mobile applications.

Acknowledgments

We are grateful to Tim Villabona of the Sensory Communication Group for his contributions in setting up the experiment at MIT and to all the subjects who participated in these experiments. We are grateful to the MIT International Science and Technology Initiatives, Empowering The Teachers programme that provided travel funds for Dr. (Mrs.) Deborah U. Ebem. The MIT researchers were supported by Grant No. R01 DC00117 from the National Institutes of Health, NIDCD.

References

1. Nilsson M, Soli SD, Sullivan JA. Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. J. Acoust. Soc. Am. 1994 Feb.95:1085–1099. [PubMed: 8132902]
2. Wong LLN, Soli SD, Liu S, et al. Development of the Mandarin Hearing in Noise Test (MHINT). Ear & Hearing. 2007 Apr.28:70S–74S. [PubMed: 17496652]
3. Wong LLN, Soli SD. Development of the Cantonese Hearing In Noise Test (CHINT). Ear & Hearing. 2005 Jun.26:276–289. [PubMed: 15937409]

4. Desloge JG, Reed CM, Braida LD, et al. Speech reception by listeners with real and simulated hearing impairment: Effects of continuous and interrupted noise. J. Acoust. Soc. Am. 2010 Jul. 128:342–359. [8]. [PubMed: 20649229]
5. Bacon SP, Opie JM, Montoya DY. The effects of hearing loss and noise masking on the masking release in temporally complex backgrounds. J. Speech Hear. Res. 1998 Jun.41:549–563.
6. George EJ, Festen JM, Houtgast T. Factors affecting masking release for speech in modulated noise for normal hearing and hearing impaired listeners. J. Acoust. Soc. Am. 2006 Oct.120:2295–2311. [PubMed: 17069325]
7. Gustafsson HA, Arlinger SD. Masking of speech by amplitude-modulated noise. J. Acoust. Soc. Am. 1994 Jan.95:518–529. [PubMed: 8120263]
8. Arlinger S, Gustafsson HA. Masking of speech by amplitude modulated noise. J. Sound Vib. 1991 Dec.151:441–445.
9. Summers V, Molis MR. Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level. J. Speech Lang. Hear. Res. 2004 Apr.47:245–256. [PubMed: 15157127]
10. Stuart A, Phillips DP. Word recognition in continuous and interrupted broadband noise by young normal hearing, older normal hearing and presbyacusic listeners. Ear & Hearing. 1996 Dec. 17:478–489. [PubMed: 8979036]
11. deLaat, JAPM.; Plomp, R. The reception threshold of interrupted speech for hearing-impaired listeners. In: Klinke, R.; Hartman, R., editors. Hearing--Physiological Bases and Psychophysics: Proceedings of the 6th international Symposium on Hearing; Springer, Berlin. 1983. p. 359-363.
12. Eisenberg LS, Dirks DD, Bell TS. Speech recognition in amplitude modulated noise of listeners with normal and listeners with impaired hearing. J. Speech Hear. Res. 1995 Feb.38:222–233. [PubMed: 7731213]
13. Ebem DU, Beerends JG, van Vugt J, et al. The Impact of Tone Language and Non-Native Language Listening on Measuring Speech Quality. J.Audio Eng. Soc. 2011 Sep.59:647–655.
14. ITU-T Recommendation P.861. Objective Quality Measurement of Telephone Band (300–3400 Hz) Speech Codecs. International Telecommunication Union; Geneva, Switzerland. 1996 Aug.
15. ITU-T Recommendation P.863. Perceptual Objective Listening Quality Assessment. International Telecommunication Union; Geneva, Switzerland. 2011 Jan.
16. Pallier C, Christophe A, Mehler J. Language-specific listening. Trends in Cognitive Science. 1997 Jul.1:129–132.
17. Lieberman, P. Intonation, Perception and Language. Cambridge, Massachusetts: M.I.T. Press; 1967.

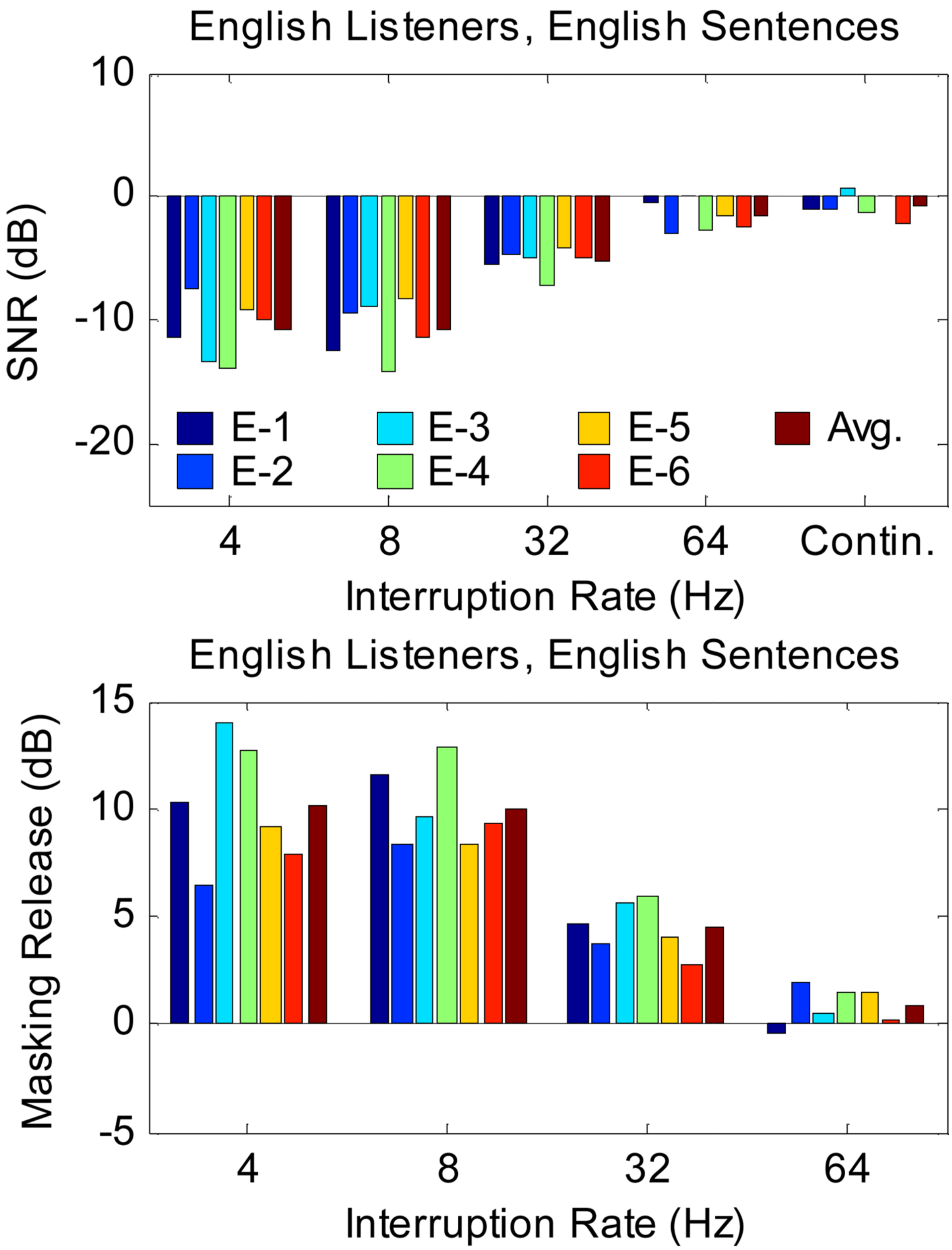

Figure 1.

English Listeners, English HINT sentences

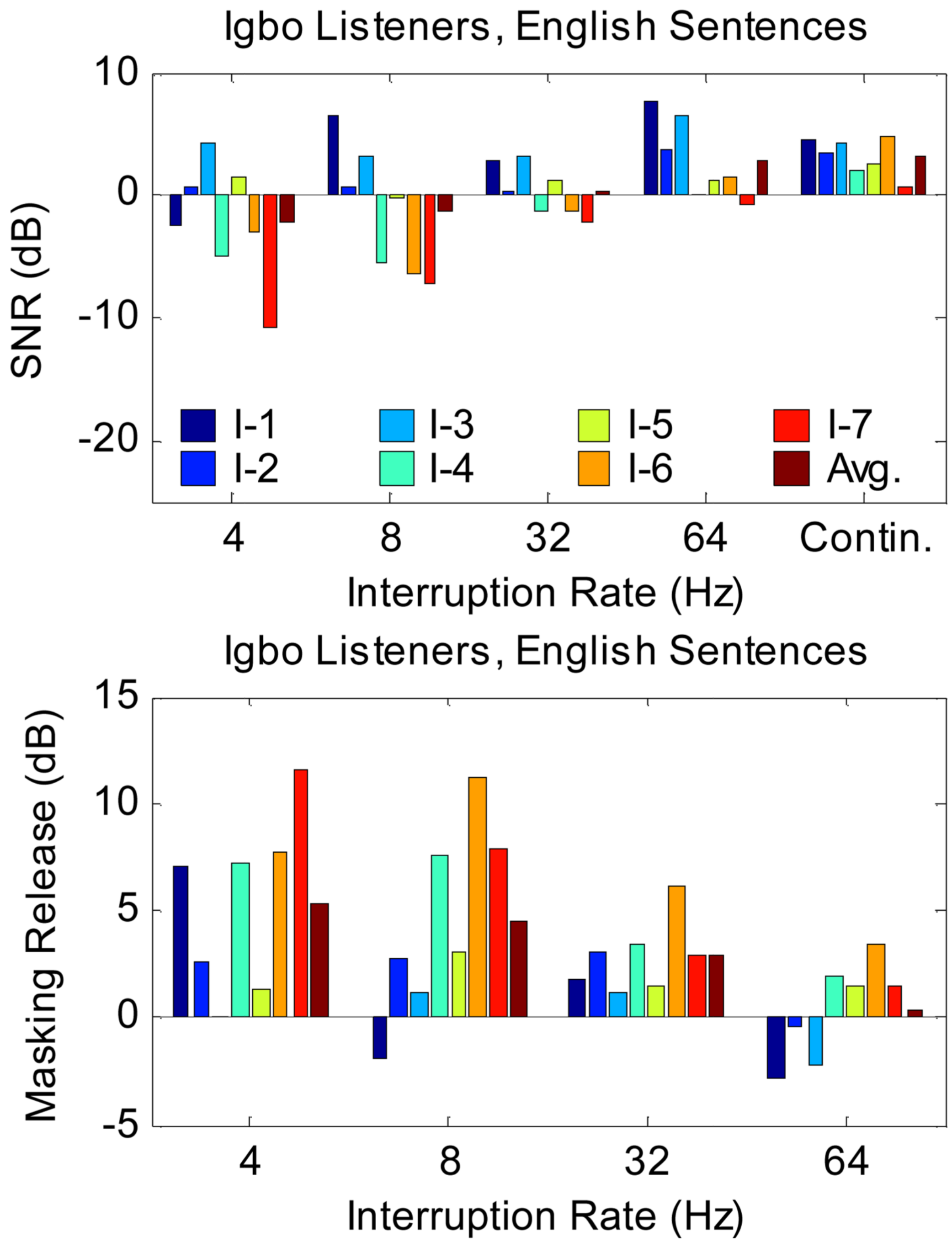

Figure 2.

Igbo listeners, English HINT sentences

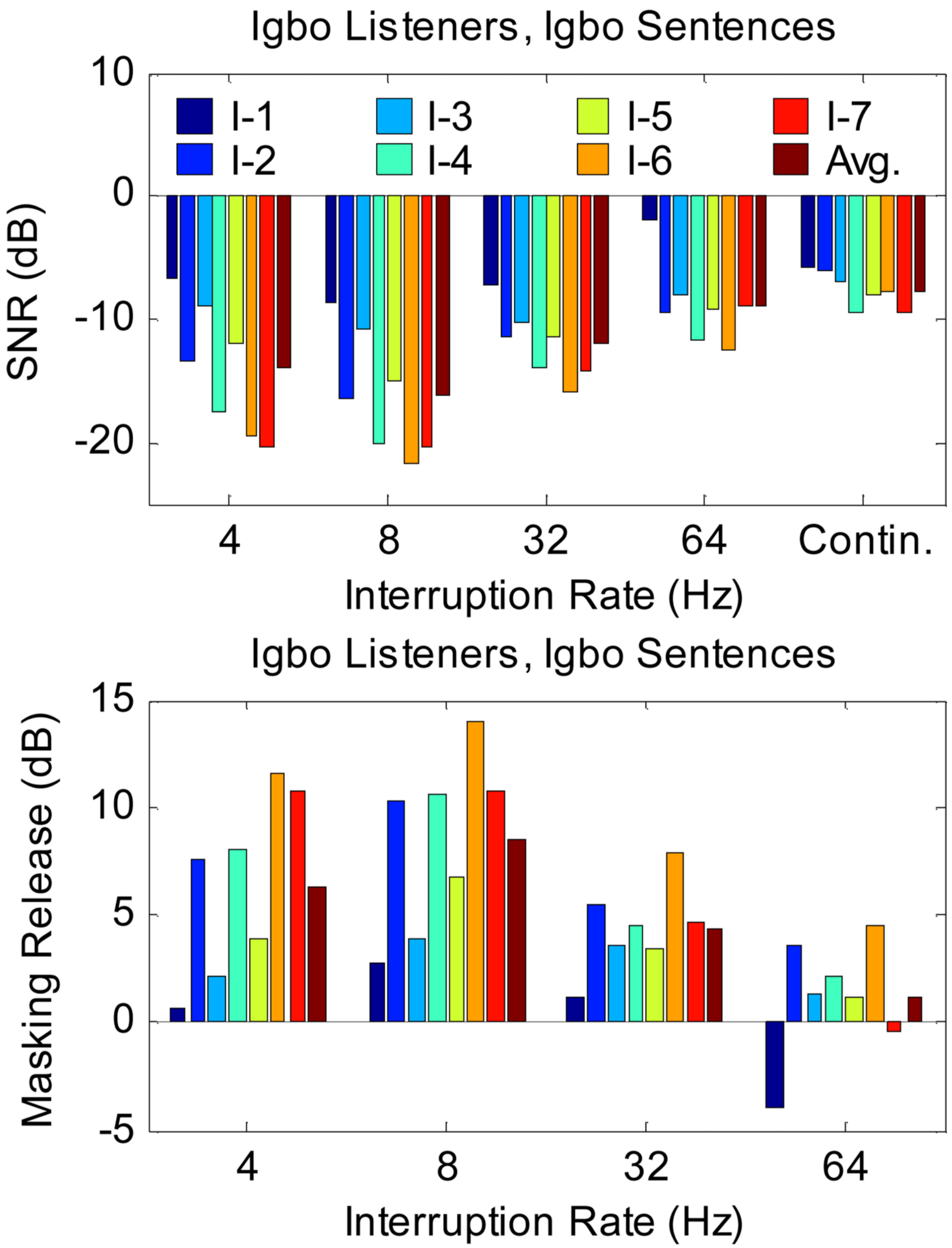

Figure 3.

Igbo listeners, Igbo sentences.

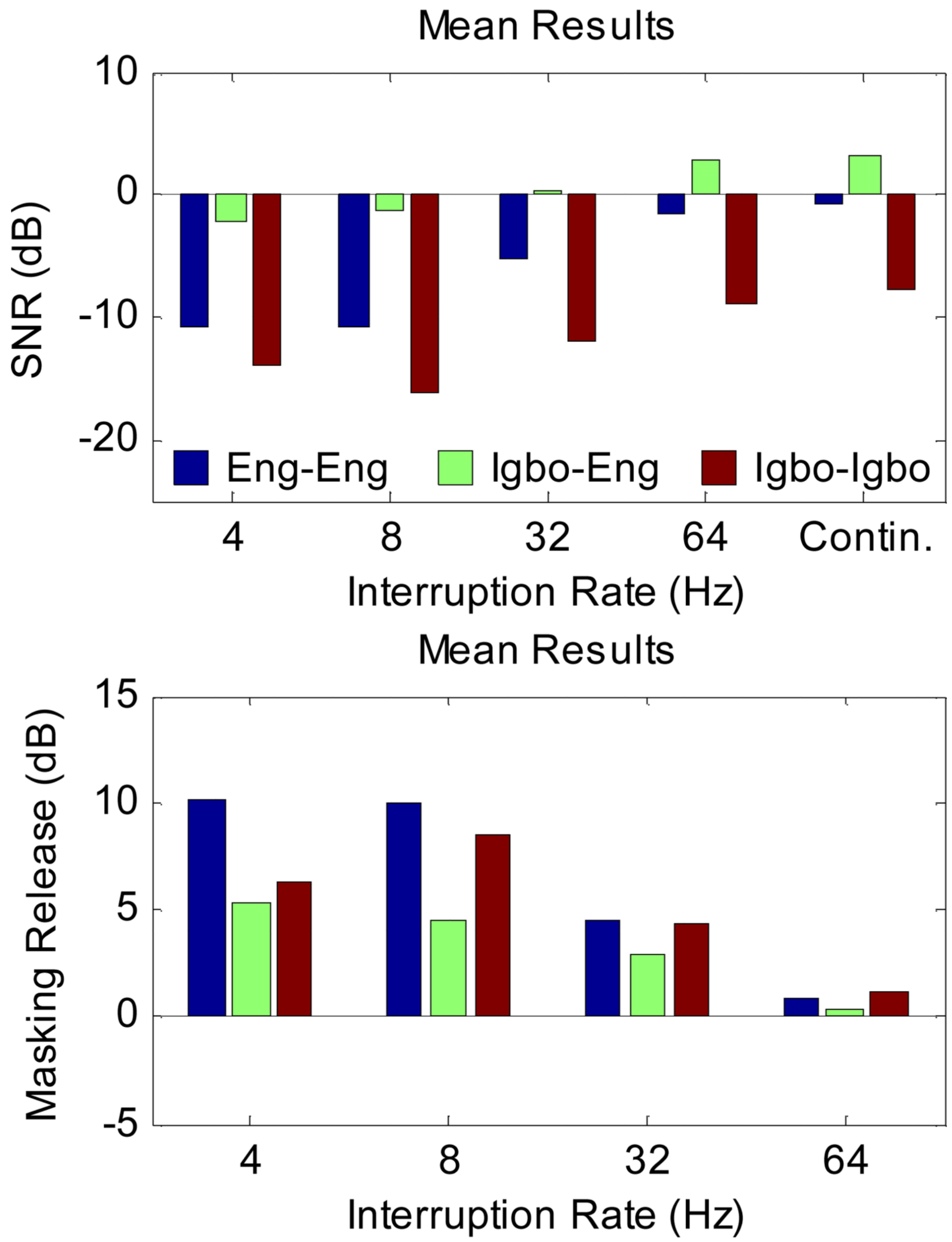

Figure 4.

Mean SNR and masking release results for the three groups: English/English, Igbo/English, and Igbo/Igbo