Published in final edited form as:
J Acoust Soc Am. 2013 October ; 134(4): . doi:10.1121/1.4820899.

Phase effects on the masking of speech by harmonic complexes: Variations with level

Tim Green[a)] and **Stuart Rosen**
UCL Speech, Hearing and Phonetic Sciences, 2, Wakefield Street, London, WC1N 1PF, United Kingdom

Abstract

Speech reception thresholds were obtained in normally hearing listeners for sentence targets masked by harmonic complexes constructed with different phase relationships. Maskers had either a constant fundamental frequency (*F*0), or had *F*0 changing over time, following a pitch contour extracted from natural speech. The median *F*0 of the target speech was very similar to that of the maskers. In experiment 1 differences in the masking produced by Schroeder positive and Schroeder negative phase complexes were small (around 1.5 dB) for moderate levels [60 dB sound pressure level (SPL)], but increased to around 6 dB for maskers at 80 dB SPL. Phase effects were typically around 1.5 dB larger for maskers that had naturally varying *F*0 contours than for maskers with constant *F*0. Experiment 2 showed that shaping the long-term spectrum of the maskers to match the target speech had no effect. Experiment 3 included additional phase relationships at moderate levels and found no effect of phase. Therefore, the phase relationship within harmonic complexes appears to have only minor effects on masking effectiveness, at least at moderate levels, and when targets and maskers are in the same *F*0 range.

I. INTRODUCTION

The degree to which target speech is masked by one or more competing voices is determined by a potentially complex interplay between factors such as energetic and informational masking (Brungart et al., 2001), the extent to which different voices can be segregated based on differences in fundamental frequency (*F*0) (Brokx and Nooteboom, 1982), and the extent to which listeners can extract information from brief “glimpses” of the target afforded by spectral and temporal fluctuations in the competing voices (Festen and Plomp, 1990; Peters *et al*., 1998). A possible approach to attempting to isolate and examine the contributions of such factors involves using maskers consisting of simplified stimuli that mimic some of the relevant features of speech. For example, Deroche and Culling (2011) found evidence suggesting an important role for masker harmonicity in *F*0-based segregation by investigating the effects of frequency modulation (FM) and reverberation on the extent of masking produced by harmonic complex tones with speech-like spectral profiles.

However, a possible complication for an approach based on using harmonic complex maskers arises from evidence that the amount of masking produced by such complexes can vary substantially according to the phase relationship between the components of the masker. Such effects have been demonstrated both for the detection of tones (e.g., Kohlrausch and Sander, 1995) and for speech recognition (Summers and Leek, 1998).

[a)]Author to whom correspondence should be addressed. tim.green@ucl.ac.uk.

PACS number(s): 43.66.Dc, 43.66.Nm [EB]

Europe PMC Funders Author Manuscripts

Summers and Leek used maskers with components summed in either positive or negative Schroeder phase (Schroeder, 1970). The resultant waveforms were time-reversed versions of each other and so had identical long-term amplitude spectra. They also had very similar, relatively flat temporal envelopes. However, the amount of masking produced in a sentence recognition task in normally hearing listeners was around 10 dB lower for positive Schroeder phase than for negative Schroeder phase.

It was suggested that this difference arose from the interaction of the different masker phase structures with the phase curvature inherent in the basilar membrane response. This results in basilar membrane waveforms that, within each cycle, either have a high-amplitude peak and a relatively long low-amplitude section (positive Schroeder phase), or a lower peak and a more similar amplitude throughout the cycle (negative Schroeder phase). The low-amplitude regions between peaks in the response to positive Schroeder phase could then allow the speech signal to be less dominated by the masker. Interestingly, no difference in masker effectiveness between Schroeder positive and negative phase was found in hearing impaired listeners. This was interpreted as indicating that nonlinear active cochlear mechanisms, differentially amplifying the low-level portions of the response in the positive Schroeder phase case, were necessary for the basilar membrane response to affect the masking of speech.

A further noteworthy aspect of Summers and Leek's (1998) data was that masker phase effects in normally hearing listeners varied with presentation level in different ways for tone detection and speech recognition. For tone detection the difference in masking between Schroeder positive and Schroeder negative phase complexes decreased as target level increased from 60 to 80 dB SPL. This is consistent with the idea that when overall level was relatively low, nonlinear active cochlear mechanisms applied greater amplification to the signal present in the low-amplitude troughs of each masker cycle in the Schroeder positive case. There is also physiological evidence that phase effects on the "peakedness" of the basilar membrane response decrease with increasing level (Summers *et al*., 2003). However, for speech recognition the opposite pattern was observed: differences between masker levels giving equivalent performance with the different phase relationships were around 8 dB for sentences presented at 60 dB SPL, and around 10 dB for 70 and 80 dB SPL presentation levels.

It was suggested that this reflected the fact that speech is a broadband signal, so that the level within any particular critical band is low, and the fact that the basilar membrane input-output function has three distinct regions (Yates, 1990). At low and high levels the function is approximately linear, while at intermediate levels it is nonlinear and compressive. Only when operating in the nonlinear region will differential gain be applied across parts of each cycle of the internal waveform and thus differences between positive and negative Schroeder phase occur. Summers and Leek (1998) suggested that in the tone detection task, the increase in presentation level tended to shift the signal up from the intermediate nonlinear region into the higher level linear region, so decreasing differences due to masker phase. On the other hand, the increase in presentation level in the sentence recognition task tended to shift the signal in individual critical bands up from the lower level, more linear, region into the nonlinear region, thus *increasing* differences due to masker phase.

Regardless of whether this explanation is correct, it is clear that presentation level can be an important determinant of the extent to which masker effectiveness is influenced by phase relationships. However, Summers and Leek's (1998) study has some important limitations with respect to considerations of speech-on-speech masking. Unlike real speech, their masker complexes had no variation in *F*0 and had equal amplitude harmonics. It is not clear to what extent masker effectiveness might be influenced by phase relationships and

presentation level for complexes that have a speech-like spectral profile and variation in *F*0. It should also be noted that Schroeder phase complexes are highly artificial and it is not clear what differences might be found with other, more natural, phase relationships.

Some relevant evidence was provided by Deroche and Culling's (2011) examination of the extent to which target and masker harmonicity affected speech reception thresholds (SRTs) in conditions in which there was a two semitone difference in *F*0 between target and masker. They used maskers consisting of harmonic complexes with components summed either in sine or random phase and filtered so as to match the spectral profile of the target sentences. Despite the fact that analysis of the maskers using simulated, level-dependent auditory filters with realistic phase responses suggested that BM responses were more peaked for the sine than the random case, SRTs were very similar across the different experiments in which the phase relationship differed. However, only a single presentation level was used, and this was towards the lower end of the levels used by Summers and Leek (1998). In addition, neither target speech nor masker complexes featured natural *F*0 variation, instead having either a constant *F*0, or a sinusoidally frequency-modulated *F*0. It is also possible that differences in the peakedness of basilar membrane responses are not as pronounced for sine compared to random phase as they are for Schroeder positive compared to Schroeder negative phase.

In the present study three experiments were carried out to address more fully the extent to which phase relationships within speech-like harmonic complex maskers affect recognition of naturally spoken sentences. The first examined the effects of masker phase relationship (Schroeder positive or Schroeder negative) on speech recognition. Masker complexes either had a constant *F*0 or dynamic variation in *F*0, similar to that seen in natural speech. For each type of *F*0 contour three presentation levels were used: a moderate level typical of speech perception experiments, and two higher levels, at which the findings of Summers and Leek (1998) suggest that effects of phase are likely to be greater. The second assessed whether the effects of phase relationships on masking at high levels differed according to the spectral profile of the masker components. The third looked for possible effects of masker phase at a moderate presentation level for a number of phase relationships beyond the highly artificial Schroeder phases used in experiments 1 and 2. These included a phase relationship that produced an approximation of a glottal voice pulse.

II. METHODS

A. Listeners

A total of 30 listeners were paid for their participation. Twelve took part in experiment 1, eight in experiment 2, and 10 in experiment 3. All spoke English as their only or primary language and had normal hearing, defined as pure-tone thresholds of 20 dB hearing level (HL) or better at octave frequencies between 500 and 8000 Hz. Ages ranged from 22 to 45.

B. Target sentences

Target speech materials were IEEE sentences (Rothauser *et al*., 1969) recorded from a male speaker of Southern British English. Each sentence contained five key words on which scoring was based. The fundamental frequency (*F*0) of the recorded sentences ranged between 93 and 151 Hz, with a median value of 115 Hz.

C. Masker complexes

Maskers were produced offline. Harmonic complexes of 30 s duration were generated with various phase relationships leading to distinct wave shapes. In experiments 1 and 2, components were in either positive Schroeder phase (SCH-P) or negative Schroeder phase

(SCH-N). Starting phase values for components in the SCH-N case were given by the formula

$$\Theta_n = -\pi n\left(n-1\right)/N, \quad (1)$$

where there are N components in total and Θ_n is the phase in radians of component n. For SCH-P complexes the initial minus sign is omitted. In experiment 3, three additional wave shapes were used. Components could have cosine phase (COS); phases that approximated a typical adult male glottal voice source (GLO) based on the Liljencrants-Fant model (Fant *et al*., 1994); or random phase (RAN). In the last case, 100 different complexes were generated, each with a different random phase relationship, and were sampled at random without replacement. Figure 1 shows waveforms and spectrograms for 50-ms sections of maskers with each type of phase relationship used, while Fig. 2 shows simulated inner hair cell (IHC) output waveforms for a channel centered at 2 kHz, derived from a recent model of the auditory periphery (Zilany *et al*., 2013). The greater peakedness resulting from the SCH-P phase relationship is clear in Fig. 2. Relatively little difference is apparent between the IHC outputs in the SCH-N, COS, and GLO cases.

In experiment 1, complexes could have either a static *F*0 or a dynamically varying *F*0. In experiments 2 and 3, all complexes had varying *F*0. In the dynamically varying case, *F*0 contours were based on passages of connected discourse from a male talker. This talker was different from the target talker but had a very similar *F*0 range (95–155 Hz, median 115 Hz). *F*0 contours were interpolated through unvoiced and silent periods using piecewise cubic Hermite interpolation in logarithmic frequency. The number of components in the complexes was set to 53 so as to ensure that components extended beyond 5 kHz for the lowest *F*0 value in the contour. Static complexes were generated with *F*0 equal to the median value of the dynamic complexes (115 Hz) and also with 53 components. Complexes were generated on a period-by-period basis, ensuring waveform continuity at the beginning and end of each cycle, which was particularly important for maskers with dynamically varying *F*0. Since *F*0 contours were based on real speech, variations in *F*0 over short time intervals were typically small. A calculation of transitional statistics showed that *F*0 typically changed very little cycle-by-cycle, with a median change of about 0.6%, which is close to the limit of discriminability (Rosen and Fourcin, 1986). Approximately 84% of adjacent cycles had a less than 2% difference in *F*0, and around 64% differed by less than 1%. Only around 1% varied by more than a semitone, and it is likely that these larger differences occurred due to the interpolation. This means that the masker complexes with varying *F*0 can be considered as periodic, as is essential for the strong pitch percept associated with most speech.

Straightforward harmonic synthesis with equal amplitude components was used for all except the GLO waveforms, for which the shape of the wave in each cycle is analytically defined. With the exception of some stimuli in experiment 2, which were left unaltered, a linear phase filter was then used to give the complexes a spectral profile corresponding to the long-term average spectrum of the target material. A sample rate of 44.1 kHz was used in generation but complexes were subsequently down-sampled to 22.05 kHz, matching the sample rate of the target sentences.

Three masker levels were used in experiment 1. For static complexes these levels set the component nearest to 2 kHz to 50, 40, or 30 dB SPL, leading to overall masker levels of approximately 80, 70, or 60 dB SPL. In experiment 2, only the highest level was used, while in experiment 3, only the lowest level was used.

Europe PMC Funders Author Manuscripts

D. Experimental variables

In experiment 1 three factors were varied: presentation level (60, 70, or 80 dB SPL), phase relationship (SCH-N or SCH-P) and type of *F*0 contour (static or dynamic). In experiment 2, phase relationship (SCH-N or SCH-P) and component amplitude (all equal or shaped to the speech spectrum) were factorially combined and masker level was fixed at 80 dB SPL. In experiment 3 masker level was fixed at 60 dB SPL and only masker phase was varied, with five relationships tested: SCH-N, SCH-P, COS, GLO, and RAN. Table I summarizes the conditions in each experiment.

E. Procedure

A randomly selected portion of the appropriate length was extracted from the 30 s of the specified masker for each trial. Target and masker were separately low-pass filtered with a 4.5 kHz cutoff frequency, before being combined and presented via Sennheiser headphones (HD650) in a sound-proof booth. Low-pass filtering used a 12th-order Butterworth filter, applied forward and backward to produce the equivalent of a 24th-order filter with zero phase lag. The onset of the target sentence was 600 ms after that of the masker complex and the masker continued for 100 ms after the offset of the target. Cosine onset and offset ramps of 100 ms were applied to the mixture. An adaptive procedure was used to estimate SRTs, defined as the signal-to-noise ratio (SNR) at which 50% of key words could be recognized correctly. SNR calculations were based on the root-mean-square level of the target and that of the masker during the period in which the target was present. In contrast to Summers and Leek (1998), the level of the masker complex, rather than that of the target speech, was fixed within a run. This approach was preferred since the effects of phase relationship were expected to vary with masker level. The first of 20 sentences (two IEEE lists) was presented at a SNR of +10 dB. SNR was decreased if more than two of the five key words were correctly identified and increased otherwise. A 10-dB change in SNR was used until the first reversal, 6.5-dB until the second reversal, and 3-dB for all subsequent reversals. SRTs were calculated as the mean of the final even number of reversals with the 3-dB step size. The number of reversals on which estimates were based ranged between 4 and 12, with a mean of 8. A single SRT estimate was obtained in each condition. The order of the conditions in each experiment was based on a randomized Latin square. For familiarization with the task, the first condition for each listener was repeated using different target sentences; the data from the first run were discarded. Within each experiment each listener was presented with the same sentences (including the familiarization run) in the same order.

III. RESULTS

A. Experiment 1: Presentation level and *F*0 contour type

Figure 3 shows SRTs for each combination of masker presentation level, *F*0 contour type and phase relationship. The most striking feature of the data was a strong interaction between masker complex phase relationship and presentation level. Performance was similar across level for SCH-N complexes but improved (SRTs were lower) with increasing level for SCH-P complexes. The difference in mean SRTs between the highest and lowest presentation levels for SCH-P complexes was 4–5 dB. For both static and dynamic *F*0 contours there was only a small effect of phase relationship on SRTs at the lowest presentation level (mean differences around 1.5 dB), but a substantial effect (5–7 dB) at the highest level.

SRTs were submitted to a three-way repeated measures analysis of variance (ANOVA) with factors of presentation level, *F*0 contour type and phase relationship. In addition to confirming that there was a significant two-way interaction between level and phase [$F(2,22) = 17.34$, $p < 0.001$], this analysis showed highly significant main effects for each

factor: level [$F(2,22) = 11.92$, $p < 0.001$], contour [$F(1,11) = 26.41$, $p < 0.001$], and phase [$F(1,11) = 86.15$, $p < 0.001$]. The three-way interaction was not significant [$F(2,22) < 1$], nor was the two-way interaction between level and contour type [$F(2,22) = 2.98$, $p = 0.072$]. There was however a significant two-way interaction between contour type and phase [$F(1,11) = 5.28$, $p = 0.042$]. Phase effects were slightly larger with dynamic contours. Averaged across presentation levels, mean SRTs with dynamic contours were around 4 dB lower for SCH-P than for SCH-N maskers (−10.4 dB and −6.5 dB, respectively), while with static contours the difference was around 3 dB (mean SRTs −11.3 dB and −8.1 dB). As these mean values show, SRTs were generally lower for static than dynamic *F*0 contours and this tendency was slightly more pronounced for maskers that were SCH-N (mean difference of 1.6 dB) than SCH-P (mean difference of 0.9 dB).

B. Experiment 2: Effects of spectral shaping

Experiment 2 examined the influence on phase effects of shaping masker complexes to match the long-term speech spectrum. Dynamic *F*0 contours and an 80 dB SPL presentation level were used—conditions which produced the largest phase effects in experiment 1. As shown in Fig. 4, performance did not differ according to whether spectral shaping was applied. With equal amplitude components, mean SRTs were respectively −4.0 dB and −10.8 dB for SCH-N and SCH-P conditions. With components shaped according to the speech spectrum the respective SRTs were −4.0 dB and −11.6 dB. A two-way repeated measures ANOVA confirmed that while there was a significant effect of phase relationship [$F(1,7) = 56.33$, $p < 0.001$], there was no significant effect of spectral shaping [$F(1,7) < 1$], and no significant interaction [$F(1,7) < 1$].

C. Experiment 3: Effects of phase relationships at moderate presentation levels

As shown in Fig. 5, there was little difference in SRTs across the different masker phase relationships at moderate presentation levels, typical of those likely to be used in speech perception experiments with normal hearing listeners. Mean SRTs ranged between −7.7 dB in the SCH-N condition and −9.2 dB in the SCH-P condition, very similar to the 1.7 dB difference observed in the equivalent conditions in experiment 1. A one-way repeated measures ANOVA showed no significant effect of phase relationship [$F(4,36) = 1.25$, $p = 0.309$].

IV. DISCUSSION

Effects of phase relationships between components were found for harmonic complex maskers that had a speech-like spectral profile and natural variation in *F*0, and thus had more in common with actual speech than those used by Summers and Leek (1998). Consistent with previous findings, there was less masking with SCH-P than with SCH-N complexes. Phase effects did not differ according to whether masker components had equal amplitude or a speech-like spectral profile. They were, however, affected by the presence of *F*0 variation, being somewhat larger for complexes with speech-like *F*0 contours than for those with a constant *F*0. Most strikingly, phase effects differed substantially according to presentation level. At the highest presentation level (approximately 80 dB SPL), mean differences between SRTs for SCH-P and SCH-N complexes were around 5–7 dB in experiments 1 and 2. At the lowest presentation level (approximately 60 dB SPL), however, differences between the two types of Schroeder phase in masking effectiveness were small, averaging around 1.5 dB in experiments 1 and 3. The outcome of experiment 3, in which additional phase relationships were examined, is consistent with Deroche and Culling's (2011) finding of no difference in the masking produced by sine and random phase complexes at a moderate presentation level.

Irrespective of any possible interaction with masker phase relationship, it might have been expected that spectral shaping of masker complexes would have led to increased SRTs relative to masking with equal amplitude components, due to a greater concentration of masker energy in spectral regions contributing most to speech understanding. The absence of such an effect here was explored by calculations of the short-time objective intelligibility measure (STOI, Taal *et al*., 2011). Since this measure cannot account for differences in masking due to phase effects, separate calculations were performed at SRTs near those observed for both positive and negative Schroeder phase in experiment 2. Although the changes in SRT for the differently shaped spectra predicted by this model were in the expected direction, they were small, always being less than about 1.5 dB. Since confidence limits for estimates of the differences in SRT for the two different spectral shapes were around +/−2 dB, it is not surprising that no significant effect of spectral shaping was found.

The incorporation of natural *F*0 variation into masker complexes led to slightly poorer performance overall, with mean differences in SRT compared to constant *F*0 complexes of around 1–1.5 dB. There was also a small but significant interaction of contour type with phase relationship, such that phase effects were around 1 dB larger with dynamic *F*0 variation. The main effect of *F*0 variation is broadly in line with Deroche and Culling's (2011) finding that masking was greater for harmonic complexes with modulated *F*0 than those with static *F*0, although in that case *F*0 variation was in the form of sinusoidal FM, rather than natural speech *F*0 contours. These results therefore provide further support to the explanation given by Deroche and Culling (2011), that *F*0 modulation interferes with the determination of periodicity in the masker complex, and so lessens the extent to which the masker can be cancelled (de Cheveigné *et al*., 1997). Presumably, the small change in SRT found here reflects the generally small short-term changes in *F*0 found in natural speech.

There was also a significant interaction of contour type with phase relationship, such that phase effects were around 1 dB larger with dynamic *F*0 variation. Note that the *F*0 of the dynamic contours went both above and below that of the static contours. It may be that phase effects are bigger at lower *F*0s, where the duration over which the phase exerts its effects is longer, and that these longer intervals outweigh the effect of the shorter intervals at higher *F*0s. This could be readily tested with static *F*0 contours at different frequencies. However, it is important to note that this effect, while significant, is small.

Comparison of the present data in constant *F*0 conditions with that obtained from normally hearing listeners by Summers and Leek (1998) reveals a similar general pattern, insofar as masking was greater for SCH-N than SCH-P complexes, phase effects increased with increasing level, and level-dependent changes occurred for SCH-P complexes, but not for SCH-N. However, there are noticeable differences across the studies in the detail of the results. Phase effects were larger (8–10 dB) in Summers and Leek (1998) than they were in the constant *F*0 conditions of the present study (1–5 dB). This may partly reflect the fact that in our experiments masker levels were fixed at similar levels to those at which target sentence levels were fixed in Summers and Leek (1998), so that the overall presentation levels were somewhat lower here. Since phase effects are smaller at lower levels this could contribute to the smaller phase effects observed here. However, since there was considerable overlap in the masker levels used across the two studies, this cannot fully account for the difference.

Other substantial methodological differences between the present study and Summers and Leek (1998) make direct comparison of outcomes somewhat difficult. For example, Summers and Leek (1998) used an unusual speech recognition procedure in which a threshold was calculated for individual target sentences. Each sentence was initially presented at a SNR of −20 dB. SNR then increased in 3 dB steps until the listener was able

to correctly identify at least three out of five key words. Since masker level was varied, this required very high initial overall presentation levels for the higher target speech levels. In addition, the repeated presentation of the same sentence may have somewhat unpredictable effects. On the one hand, the listener is able to accumulate information over different presentations of the sentence, which might allow the criterion level of key word identification to be achieved at a lower SNR than in a procedure in which each sentence is presented only once. On the other hand, our own experience with the adaptive procedure devised by Plomp and Mimpen (1979), in which the initial sentence in a run is repeated with an increasing SNR until correctly identified, suggests that it may sometimes be difficult for listeners to overcome the influence of initially misperceiving one or more words within the sentence, even if they know that their initial response is mistaken. This tendency to continue perceiving particular words incorrectly could tend to inflate SRT estimates in Summers and Leek's (1998) procedure, compared to a more typical adaptive procedure.

A further potentially important difference is that the target speech in Summers and Leek's study came from a female talker, whose mean *F*0, while not specified, was presumably considerably higher than the 100 Hz *F*0 of their maskers. In contrast, in the present study the median *F*0 of the male target talker was the same as that of the masker. Assuming that phase effects on masking are attributable to SCH-P complexes producing a more peaked internal response, it is possible that such effects will be greater when the *F*0 of the target speech is substantially higher than that of the masker. The low-amplitude section of the response to a cycle of a SCH-P masker will contain only part of a target pitch period when the target *F*0 is similar to that of the masker complex, but may contain one or more complete pitch periods for higher target *F*0s. This more complete representation of periodicity may facilitate the extraction of the acoustic structure of the target speech and so increase the differences between SCH-P and SCH-N maskers.

It could be expected that a larger difference in *F*0 between target and masker would tend to lead to better overall performance, irrespective of any contribution of phase effects. However, SRTs for SCH-N maskers in the present study were around −8 dB regardless of presentation level. In contrast, Summers and Leek's Fig. 5 shows that SRTs for SCH-N maskers were approximately 0 dB. Other procedural differences described above may have contributed to this difference and the inherent intelligibility of the target talkers may have differed. Nonetheless, the SCH-N performance of Summers and Leek's normally hearing listeners does seem rather poor, and it is noteworthy that it did not differ from that of their hearing impaired listeners in the equivalent condition.

The goal of the present study was to examine the effects of phase relationships on masking by harmonic complexes in conditions with more in common with typical speech perception experiments than those employed by Summers and Leek (1998). The largest influence on phase effects was presentation level and it is possible that this factor is primarily responsible for the absence of phase effects in Deroche and Culling (2011). Only relatively small phase effects (2.7 dB for constant *F*0 and 3.3 dB for speech-like *F*0 variation) were observed at 70 dB SPL, which was very similar to the level used by Deroche and Culling (2011). There may also have been contributions from the fact that *F*0 variation was not natural in that study for either target speech or masker complexes, and that random and sine phases, rather than SCH-N and SCH-P were compared.

The present study has demonstrated that incorporating speech-like spectral profiles and natural *F*0 variation into complex harmonic maskers does not eliminate the possibility of phase effects on the extent of masking of natural speech. However, such effects appear to be highly level dependent and, at least in conditions where target and masker *F*0 were similar, were substantial only for presentation levels considerably higher than those typically used in

speech perception experiments with normally hearing listeners. Research with hearing-impaired listeners would, of course, likely require higher presentation levels. However, the results of Summers and Leek (1998) make it clear that phase effects are unlikely to occur in such listeners.

Acknowledgments

This work was supported by the MRC (UK) Grant No. G1001255. Much of the testing was carried out by Hannah Williams. We are grateful to John Culling for initially highlighting the issue of phase relationships in harmonic complexes to us, and for further helpful discussions. Thanks also to Ian Bruce, Torsten Dau, Claus Elberling, Gaston Hilkhuysen, Filip Munch Rønne, and Van Summers for various useful pieces of information, stimuli and software; to Muhammad Zilany whose software was used to generate the auditory model output waves in Fig. 2; to Alan O Cinneide whose software was used to generate the GLO stimuli; to Cees Taal whose software was used for calculations of the STOI; and to Sam Eaton-Rosen for assistance in generating Figs. 1 and 2.

References

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

Brungart DS, Simpson BD, Ericson MA, Scott KR. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Am. 2001; 110:2527–2538. [PubMed: 11757942]

de Cheveigné A, McAdams S, Marin CMH. Concurrent vowel identification. II. Effects of phase, harmonicity, and task. J. Acoust. Soc. Am. 1997; 101:2848–2856.

Deroche MLD, Culling JF. Voice segregation by difference in fundamental frequency: Evidence for harmonic cancellation. J. Acoust. Soc. Am. 2011; 130:2855–2865. [PubMed: 22087914]

Fant, G.; Kruckenberg, A.; Liljencrants, J.; Båvegård, M. Voice source parameters in continuous speech. Transformation of LF-parameters; Proceedings of the ICSLP-94; Yokohama. 1994; p. 1451-1454.Vol. 3

Festen JM, Plomp R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. J. Acoust. Soc. Am. 1990; 88:1725–1736. [PubMed: 2262629]

Kohlrausch A, Sander A. Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. J. Acoust. Soc. Am. 1995; 97:1817–1829. [PubMed: 7699163]

Peters RW, Moore BCJ, Baer T. Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. J. Acoust. Soc. Am. 1998; 103:577–587. [PubMed: 9440343]

Plomp R, Mimpen AM. Speech-reception threshold for sentences as a function of age and noise-level. J. Acoust. Soc. Am. 1979; 66:1333–1342. [PubMed: 500971]

Rosen, S.; Fourcin, AJ. Frequency selectivity and the perception of speech. In: Moore, BCJ., editor. Frequency Selectivity in Hearing. Academic; London: 1986. p. 373-487.

Rothauser EH, Chapman ND, Guttman N, Nordby KS, Silbiger HR, Urbanek GE, Weinstock M. Recommended practice for speech quality measurements. IEEE Trans. Audio Electroacoust. 1969; 17:225–246.

Schroeder MR. Synthesis of low-peak-factor signals and binary sequences with lol autocorrelation. IEEE Trans. Info. Theory. 1970; 16:85–89.

Summers V, de Boer E, Nuttall AL. Basilar-membrane responses to multicomponent (Schroeder-phase) signals: Understanding intensity effects. J. Acoust. Soc. Am. 2003; 114:294–306. [PubMed: 12880042]

Summers V, Leek MR. Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners. Hear. Res. 1998; 118:139–150. [PubMed: 9606069]

Taal CH, Hendriks RC, Heusdens R, Jensen J. An algorithm for intelligibility prediction of time-frequency weighted noisy speech. IEEE Trans. Audio Speech Lang. Process. 2011; 19:2125–2136.

Europe PMC Funders Author Manuscripts

Yates GK. Basilar membrane nonlinearity and its influence on auditory-nerve rate-intensity functions. Hear. Res. 1990; 50:145–162. [PubMed: 2076968]

Zilany MSA, Bruce IC, Ibrahim RA, Carney LH. Improved parameters and expanded simulation options for a model of the auditory periphery. Assoc. Res. Otolaryngol. Abstr. 2013; 36:440.

Europe PMC Funders Author Manuscripts

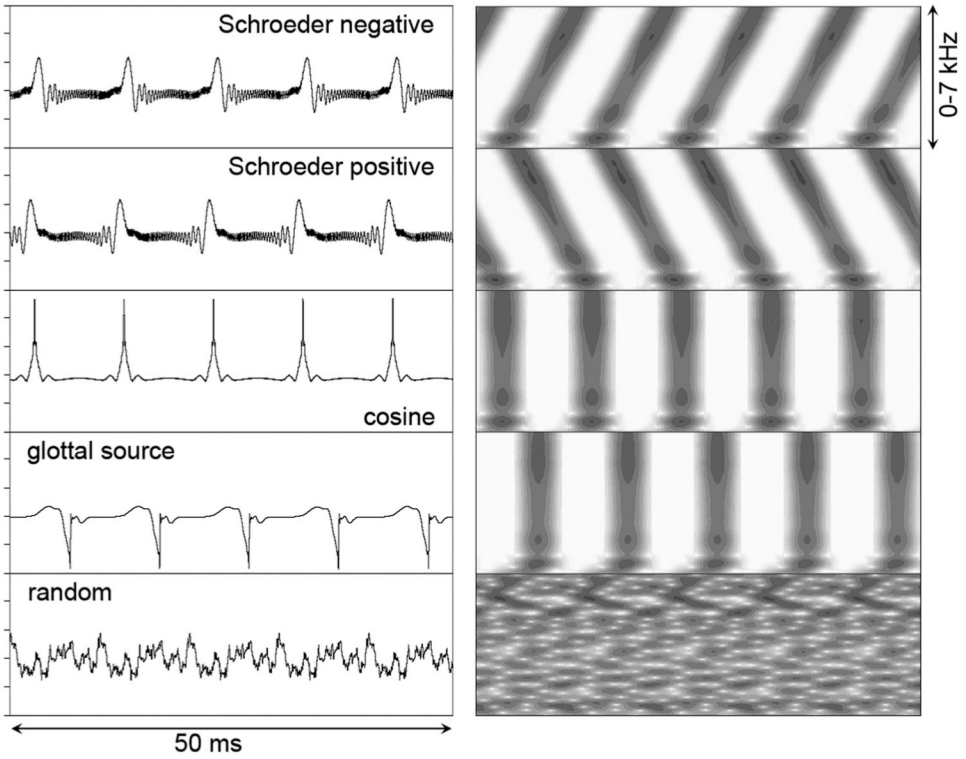

FIG. 1.

Waveforms (left) and wideband spectrograms (right) of 50-ms sections of maskers with the different phase relationships used. The range of voltages is the same for all five waveforms and is on an arbitrary linear scale. All complexes shown were shaped to match the long-term average spectrum of the speech targets. Low-pass filtering at 4.5 kHz was applied at run time but is not reflected in these representations. As is typical for spectrograms, an equalizing filter was applied to the original waveforms in order to “whiten” their spectra. The random-phase wave shown had the median peak factor of a set of 100 generated waves.

Europe PMC Funders Author Manuscripts

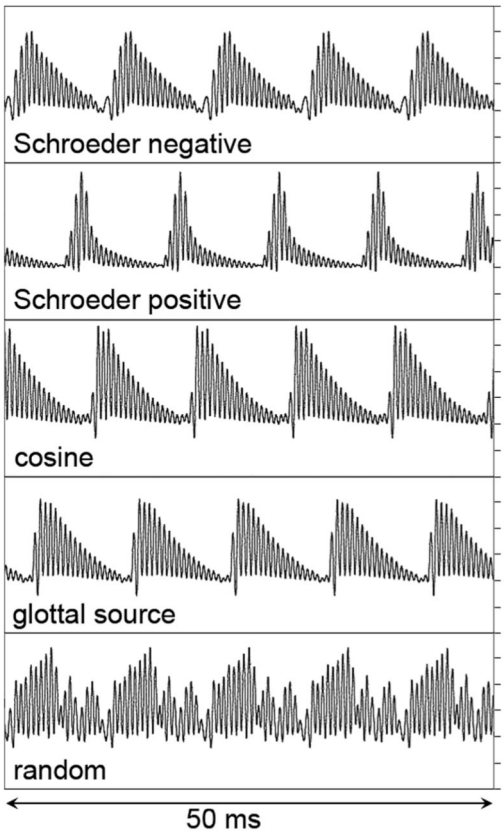

FIG. 2.
Simulated inner hair cell output waveforms from a 2-kHz filter derived from the Zilany *et al*. (2013) model of the auditory periphery for the same maskers shown in Fig. 1. The range of voltages is the same for all five waveforms and is on an arbitrary linear scale.

Europe PMC Funders Author Manuscripts

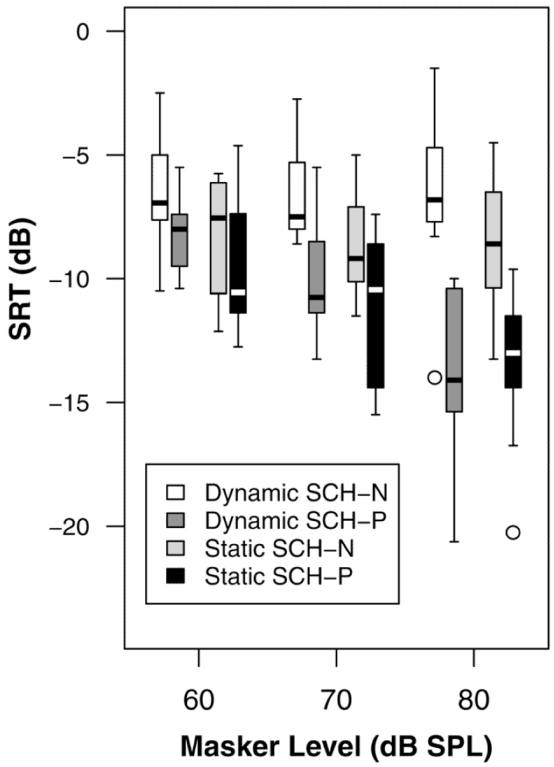

FIG. 3.

Box plots of SRTs for each combination of *F*0 type (dynamic or static) and phase relationship (Schroeder positive or Schroeder negative) for each of the three presentation levels in experiment 1. All maskers had components shaped according to the speech spectrum. The bar within each box shows the median, the extremes of the box show the first and third quartiles, whiskers extend to the most extreme data point no more than 1.5 times the interquartile range from the box, points outside that range are shown by open circles.

Europe PMC Funders Author Manuscripts

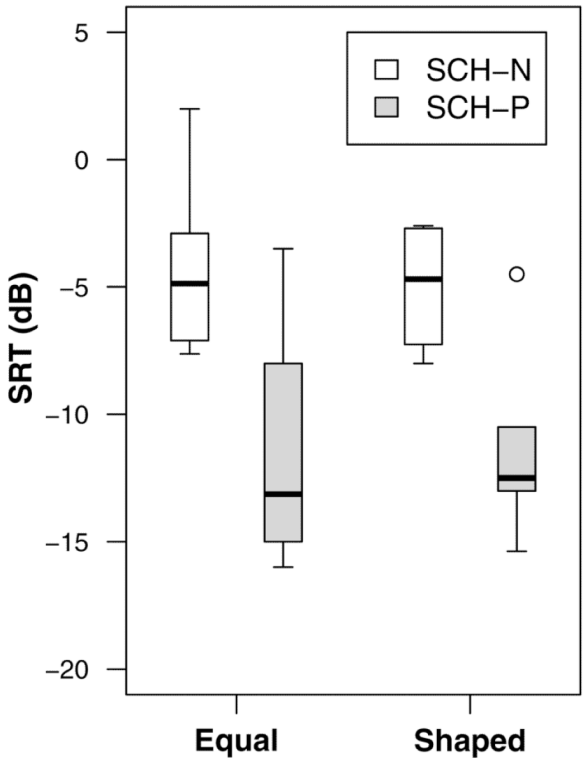

FIG. 4.

Box plots of SRTs for Schroeder positive and Schroeder negative complexes with either equal amplitude components or components shaped according to the speech spectrum in experiment 2. Presentation level was 80 dB SPL. All maskers had dynamically varying *F*0.

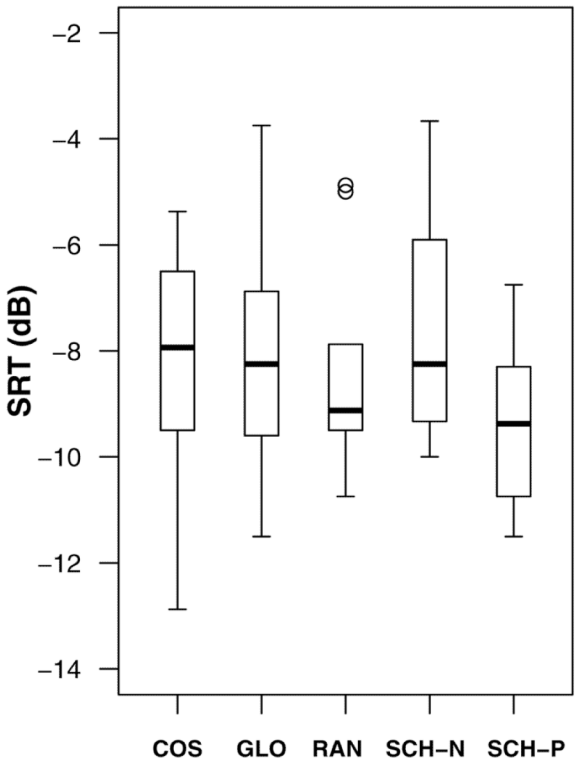

FIG. 5.
Box plots of SRTs for masker complexes with various phase relationships in experiment 3. Presentation level was 60 dB SPL. All maskers had dynamically varying *F*0 and components shaped to match the spectrum of the target speech.

Europe PMC Funders Author Manuscripts

TABLE I

Summary of conditions in each of the three experiments

Expt	Masker phases	Masker levels (dB SPL)	Component amplitude	*F*0 Contour	Number of conditions
1	SCH-N	60	Shaped	Dynamic	12
	SCH-P	70		Static	
		80			
2	SCH-N	80	Shaped	Dynamic	4
	SCH-P		Equal		
3	SCH-N	60	Shaped	Dynamic	5
	SCH-P				
	COS				
	GLO				
	RAN				