Language and Cognition (2025), 17, e71, 1–20
doi:10.1017/langcog.2025.10030

ARTICLE

I'm whispering a white Christmas: masking relations in hallucinatory speech

Mark Scott[1] and Tommi Tsz-Cheung Leung[2]

[1]Department of English Literature and Linguistics, Qatar University, Doha, Qatar and [2]Department of Cognitive Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates
Corresponding author: Tommi Tsz-Cheung Leung; Email: leung@uaeu.ac.ae

(Received 16 February 2025; Revised 02 July 2025; Accepted 22 August 2025)

Abstract
Auditory verbal hallucinations are a common phenomenon in the general population, with many people without psychological issues reporting the experience. In the 'White Christmas' method to induce auditory hallucinations, participants are told that they will be played a portion of the song 'White Christmas' and are asked to report when they hear it. Participants are presented only with stochastic noise; still, a large proportion of participants report hearing the song. The experiments reported here investigate how masking relationships modulate verbal hallucinations in the White-Christmas effect. Specifically, we tested how the effect is modulated by different kinds of maskers (multi-talker babble versus spectrally matched speech-shaped stochastic noise) and different kinds of expectation of the speech being masked (expecting a 'normal' modal voice versus a whispered voice behind the masking). The White Christmas effect was replicated, and the rate of verbal hallucinations was higher for multi-talker babble than for spectrally-matched speech-shaped stochastic noise. In addition, a trend for a higher rate of hallucination for whispered voices was found. These results confirm the role of masking relations in the White Christmas effect and reinforce the similarity between the White Christmas effect and continuity illusions such as phoneme restoration.

Keywords: auditory verbal hallucination; perception; phonetics; speech

1. Introduction

A common technique for eliciting auditory verbal hallucinations (AVH) is to present people with a masking sound and suggest to them the presence of a target sound under the masker. The most famous demonstration of this (using music rather than plain speech) is the *White Christmas effect* in which participants are played stochastic ('white') noise and told to indicate when they hear a snippet of the Bing Crosby song 'White Christmas' underneath the noise (hence the name of the effect). Many participants report hearing the song, despite there being nothing in the audio signal but stochastic noise (Barber & Calverley, 1964). Several replications have demonstrated the effect with

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

speech as well (e.g., Feelgood & Rantzen, 1994; Fernyhough et al., 2007; Hartley et al., 2017). Several modulators of the effect have been explored, including caffeine (Crowe et al., 2011), fantasy proneness (Merckelbach & van de Ven, 2001), hypnosis (Barber & Calverley, 1964), cognitive suppression (Rassin & Heiden, 2006) and stress (Hoskin et al., 2014). The White Christmas effect is structurally similar to phoneme restoration (discussed below) in which people hear a deleted speech sound as present when the sound's empty position is masked. Based on this similarity, and the dependence of phoneme restoration on masking quality, the current experiments explore whether similar masking relationships modulate the White Christmas effect.

AVH have been induced using methods different from, but resembling, the White Christmas effect. For example, B.F. Skinner presented participants with a recording of overlapping indistinct vowel sounds, leading to reports of hearing words (discussed in Deutsch, 2019). Young et al. (1987) asked participants to simply imagine sounds (and visuals). Many participants reported having vivid sensory experiences, though none reported believing the experiences to be genuine percepts and so this is not clearly a demonstration of hallucination. In a signal-detection set-up very similar to the White-Christmas methodology, Frost et al. (1988) played recordings of words masked with noise modified to have the same amplitude profile as the words. This masked audio-stimulus was presented alongside text of the word, leading participants to report hearing the word more vividly. In another approach, Feelgood and Rantzen (1994) were able to induce AVH in non-clinical[1] participants by having them listen to spliced together backward recordings of a male voice. There is also the similar *verbal transformation effect* in which simply playing the same word or phrase on a loop leads a hearer, after sufficient repetition, to start hearing the loop as containing different words (Reisberg, 1989; Warren & Gregory, 1958). Diana Deutsch (2003) has produced a version of this illusion, the methods for which are available for download at http://philomel.com/phantom_words/pages.php?i=1010. In this version, a two-word (or two-syllable) sequence is played from speakers positioned to the left and right of the listener, with a slight temporal offset. The listener will typically begin to hear words other than those being played.

1.1. Auditory verbal hallucinations

AVH, or hearing voices, is most commonly associated with schizophrenia, and is indeed one of the key diagnostic symptoms of the disease (Frith, 1992). However, hearing voices is not limited to people with schizophrenia, and is, perhaps surprisingly, a common experience in the general population, with approximately 9.6% of people with no known psychological issues reporting the experience (van Slobbe-Maijer, 2019).[2] The *Diagnostic and statistical manual of mental disorders* (2013, 23) provides a definition of hallucination:

> Hallucinations are perception-like experiences that occur without an external stimulus. They are vivid and clear, with the full force and impact of normal perceptions, and not under voluntary control.

[1]Here and elsewhere "non-clinical" is used to mean people with no known mental illness.

[2]Perhaps the most familiar experience that relates to this is the "Phantom Phone Vibration" many people experience, in which they hallucinate that their cellphone has vibrated (Drouin et al., 2012).

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

To account for the fact that many hallucinations are more likely to occur in the presence of certain sensory stimuli, Slade and Bentall (1988, 23), have proposed a broader definition:

> Any percept-like experience which (a) occurs in the absence of an appropriate stimulus, (b) has the full force or impact of the corresponding actual (real) perception and (c) is not amenable to direct and voluntary control by the experiencer.

This definition would include the sounds heard in the White Christmas effect as hallucinations, since, while there is an auditory stimulus inducing the experience, there is not an *appropriate* auditory stimulus. Under this definition, however, the distinction between hallucination and illusion becomes blurred. An illusion is usually considered a non-veridical interpretation of a sensory stimulus, rather than a perception in the absence of a sensory stimulus (e.g., Hoffman, 2012). In the White-Christmas effect, participants listen for sounds buried under a masking sound, and so the effect could be interpreted more along the lines of an illusion than a hallucination (i.e., *misinterpreting* components of the masking sound as being speech or music). Under this interpretation, the White-Christmas effect is an illusion very similar to the phoneme restoration effect, in which people hear a speech sound that has been artificially removed from a recording and replaced with a masking sound, such as a cough (Samuel, 1996; Warren, 1970). People generally do not notice that a phoneme has been replaced and, when asked, are usually unable to determine at what point in the word the masking sound occurred. It is generally assumed that the lexical and phonological context in which the missing phoneme is situated provides an expectation of hearing the particular phoneme and that this expectation allows the missing sound to be heard under the masking sound (Srinivasan & Wang, 2005). There is no sharp boundary between illusion and hallucination, however, as no sensory experience happens without a context. Even in a sensory deprivation situation there is background sensory noise internal to the body itself that provides potential for sensory misinterpretation. We are not particularly concerned about whether the White Christmas effect should be categorized on one side or the other of the hallucination/illusion divide; however, since the term 'hallucination' has generally been used in the White Christmas literature, we will retain that label here for the sake of consistency.

There are two primary theories about the mechanism of AVH. One argues that the voices are the result of a misattribution of an internal mental event, such as a memory or a person's own inner speech (e.g., Bentall et al., 1991; Brookwell et al., 2013; Frith, 1992). Hugdahl has proposed a variation on this model which assumes that the mental event is perceptual and explains misattribution as the result of a failure of inhibitory control (2009). The other approach suggests that voice hearing is the result of a signal detection bias in which the hearer's own expectations lead to the misinterpretation of an external signal (e.g., Alganami et al., 2017; Bentall & Slade, 1985). This explanation of AVH, as with the White Christmas effect, blurs the line between hallucination and illusion as the 'hallucination' is being caused by misinterpretation of an external sensory signal (rather than in the absence of a sensory signal). These two potential mechanisms for explaining AVH are not mutually exclusive, as the biasing top-down expectation in the signal-detection viewpoint is often assumed to be an internal mental event (Moseley et al., 2016). This literature is reviewed in Toh et al. (2022).

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

The signal-detection theory of AVH would seem to be an obvious candidate for explaining the White Christmas effect, given that the White Christmas effect relies on biasing the interpretation of a masking sound to hear a (missing) signal. However, it is plausible that it is a mental event (such as a memory of a word or the word being spoken in inner speech) that provides the top-down expectation that influences processing of the masking sound in the White Christmas effect. Given this, both theories on the mechanism of AVH may be relevant to the White Christmas effect. It is possible that there are in fact multiple mechanisms that trigger voice hearing and thus different *types* of voice-hearing experience. This is suggested by the fact that the phenomenology of voice hearing is quite diverse – some people hear a single voice, while others hear multiple voices; some people hear the voice as located in external space, while others hear voices that are experienced internally (Toh et al., 2020).

The White Christmas effect induced in the experiments reported below is with people not suffering from any known mental illness (which is how the White Christmas effect is normally used). The question remains as to whether the voice hearing of non-clinical populations can be informative about clinical populations. As reviewed in Toh et al. (2022), the *discontinuous* model argues that the voices associated with psychosis have a different etiology from those in non-clinical populations, while the *continuum* hypothesis argues that these experiences are essentially the same and differ simply in degree. The current experiment does not address this debate. However, if the continuum hypothesis proves to be correct, then research on the White Christmas effect in non-clinical populations will obviously be relevant to research on hallucinations related to psychosis.

1.2. Phoneme restoration

If viewed as being the result of signal detection bias, the White Christmas effect is very similar to the phoneme restoration effect in which people hear a speech sound that has been artificially removed from a recording and replaced with a masking sound, often stochastic noise (Samuel, 1996; Warren, 1970). People generally do not notice that a speech sound has been replaced, and when asked, are usually unable to determine at what point in the word the masking sound occurred. It is generally assumed that the lexical and phonological context in which the missing phoneme is situated provide an expectation of hearing the particular phoneme and that this expectation allows the missing sound to be heard under the masking sound (e.g., Bashford & Warren, 1987; Srinivasan & Wang, 2005; Trout & Poser, 1990). This is a case of expectation causing noise (or other masking sounds) to be heard as speech and so is clearly similar to the White Christmas effect.

Given that the phoneme restoration effect is modulated by masking relations, the experiments reported here examine how masking relations may also modulate the White Christmas effect. The probability of inducing phoneme restoration is dependent on the masking potential rule:

> The necessary condition for the auditory continuity illusion is that the acoustic (spectral, temporal, and spatial) characteristics of the interrupting sound must be sufficient to mask the interrupted sound if the two sounds were presented simultaneously. (Kashino, 2006, 319–320)

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

As elaborated by many researchers (e.g., Bashford et al., 1992; Bashford & Warren, 1987; Warren et al., 1972), this principle seems to be of broad application and in its broader form is often referred to as 'auditory induction'. For example, in a non-speech parallel to phoneme restoration, an intermittent tone is heard as continuous if the silent gaps in the tone are filled with noise that *would be* sufficient to mask the tone (Bregman, 1990).

Despite the similarity between phoneme restoration and the White Christmas effect there has been little research on how the White Christmas effect is modulated by masking relations. The current experiments address this gap by examining two types of masking sound (multi-talker babble versus spectrally matched speech-shaped noise) and two types of phonetic expectation (expecting to hear a 'normal' modal voice versus expecting to hear a whispered voice). **Experiment One** examines the 'masking' side of the White Christmas effect, looking at how babble versus stochastic noise modulates the effect. **Experiment Two** examines these two maskers in the context of an expectation of an easily masked kind of speech (whisper). Of course, as there is no actual signal in the White Christmas effect, there is no actual masking either, there is only *potential* masking.

2. Experiment one

This experiment explores the effect of (potential) masking on the White Christmas effect. The prediction is that there will be more AVH for masking from babble than for masking from stochastic noise. This prediction is based on the similarity between the White Christmas effect and phoneme restoration. Given this similarity, it is plausible that similar principles will govern how the quality of the masker influences the likelihood of the effect.

When it comes to the White Christmas effect, there is no actual speech to be masked, and so the masking potential rule cannot be applied directly as it can in phoneme restoration where the surrounding speech provides a reference point for masking requirements of the deleted sounds of the speech signal. However, if the same principle is at work, then the White Christmas effect should still be influenced by the characteristics of the masker. Specifically, a higher rate of White Christmas effect should occur when the masker is multi-talker babble than when the masker is speech-shaped stochastic noise, as babble is generally the better masker of speech (Simpson & Cooke, 2005).

The primary reason that babble is a more effective masker is that it has fragments of harmonic content. This harmonic content is a particularly effective masker of the harmonic components of speech (the voiced components) (Alm et al., 2009; Popham et al., 2018; Steinmetzger & Rosen, 2015). This is seen in phoneme restoration, where voiced consonants are less likely to be restored than voiceless consonants when the masking used is stochastic noise (Kim & Davis, 2007; Trout & Poser, 1990). It should be pointed out that the greater masking of babble over stochastic noise does not hold when the number of masking speakers in the babble is small enough that there are 'gaps' in the masker allowing the perceiver to 'glimpse' the target (Darwin, 2008). It should also be pointed out that periodicity in the masker can actually reduce masking when there is a single masking voice, that is, when the task is to distinguish a target voice from a single other voice (de Cheveigné et al., 1995). In this situation, the perceptual system can use the periodicity of the masker to segregate it from the

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

periodicity of the target (as long as they are sufficiently different in fundamental frequency).

In addition to this *harmonic content* on *harmonic content* masking, there are other factors that make babble a more effective masker of speech than stochastic noise. Simpson and Cooke (2005) point out that the auditory system is particularly sensitive to sound onsets, and multi-talker babble contains many more speech onsets than speech-shaped noise, and further, that real speech (as contained in babble) more naturally causes the auditory system to pay attention to it than does noise. These factors are related to *informational masking* rather than the *energetic masking* that is typically considered in phoneme restoration and other forms of auditory induction. This issue is taken up in the General Discussion below.

To test the dependence of the White Christmas effect on masker type, **Experiment One** induced the White Christmas effect using multi-talker babble and spectrally matched speech-shaped stochastic noise. As a preview of results, there was a higher rate of AVH for the multi-talker babble in comparison to the spectrally matched speech-shaped stochastic noise.

2.1. Methods

2.1.1. Participants

There were 42 participants recruited from United Arab Emirates and Qatar University. All were native Arabic-speaking and female.[3] All participants reported no known hearing problems. The average age was 19.76 years (SD = 1.74). Participants were paid or given course credit for their participation.

2.1.2. Ethics approval

Both experiments reported in this paper were run with the approval of the respective university ethics boards; Ethics approval numbers: United Arab Emirates University ERSC_2023_2569 and Qatar University QU-IRB 990-E18.

2.1.3. Procedure

Participants sat in front of a computer screen wearing headphones (Extreme Isolation 30). In each trial, the participant was presented with a written target word. Given the demography of participants, all visual and audio stimuli were presented in Arabic.

The participants were told to listen for the target word during the masker that was to follow. Any time they heard the full word (not just a part of it), they were to hit the spacebar on the computer keyboard. They were told that the word could appear multiple times during a trial, so they could hit the spacebar more than once per trial. They were also warned that the word may not occur at all during the trial (or throughout the experiment), so it was fine not to hit the spacebar at all for any given trial, or indeed over the entire duration of the experiment.

When ready to start a trial the participant pressed the computer's 'enter' key at which point either multi-talker babble or spectrally-matched speech-shaped stochastic noise would play for 35 seconds (the sound started at silence and rose to

[3]The departments from which participants were recruited (for both **Experiment One** and **Two**) only have female students and so a sex-balanced sample was not possible.

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

approximately 80 dB over the first 2 seconds, remained at that intensity for 31 seconds and then ramped back down to silence over the last 2 seconds). So that participants would have something to look at while listening to either of the two forms of masking sound, a 100-dot random dot kinematogram was presented on the screen for the duration of the masking sound. The kinematogram was added after initial piloting because several participants complained that staring at a blank screen for the 35 seconds of the trial made it difficult to remain on task.

The experiment was run by research assistants who were familiar with the White Christmas effect but were not aware of the predicted effects. They were instructed to assure participants that many people never hear the specified sound and so not to feel that they had to hit the button (that it was fine never to hit the button throughout the experiment).

A schematic outline of the experimental set-up is shown in Figure 1, and a schematic outline of a single trial is shown in Figure 2.

There were 14 such trials over the course of the experiment. The entire experiment, including instructions and initial familiarization and practice, lasted about 20 minutes. The dependent measure was the average number of reported instances of

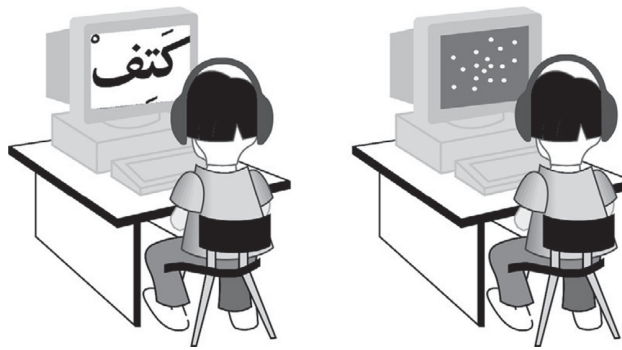


Figure 1. Schematic of experiment set-up.

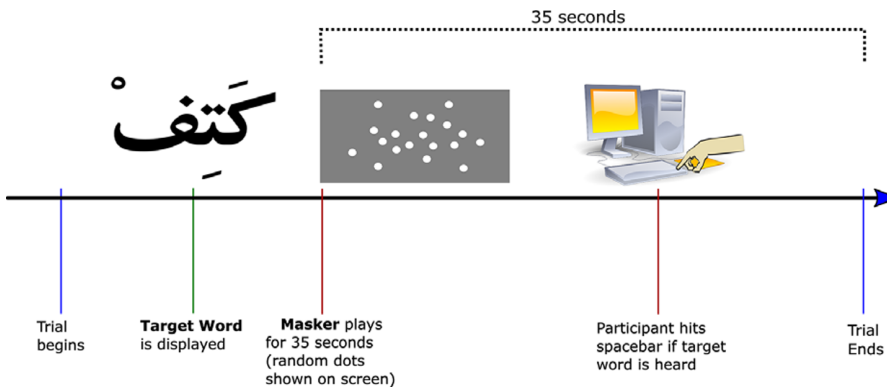


Figure 2. Schematic of a single trial.

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

hearing the target word during a trial (AVH per trial). For each participant, this was measured simply as the average number of times that participant hit spacebar on each trial.

2.1.4. Materials

Masking sounds The auditory stimuli for this experiment were two types of masking sound: Multi-talker babble and spectrally matched speech-shaped stochastic noise. The multi-talker babble was created by downloading audio files from the Qatari Arabic Corpus, available at: http://www.isle.illinois.edu/dialect/data.shtml.

This corpus consists of interviews from the Al-Jazeera network (207 minutes of relatively formal Arabic from a variety of Arabic dialects, recorded in Qatar) and 110 minutes of 'Sabah El-Doha' an interview and news broadcast with several Arabic dialects used, recorded in Qatar using relatively informal Arabic. Twenty-one of these files (or parts of files) were selected that had no music or other non-speech sounds. These selections were all normalized to the same intensity. The waveforms were cut at crossing ('zero') points to avoid pops where waveforms were joined together. These sounds were then overlapped to create a multi-talker babble of approximately 30 voices. While individual words could not be distinguished in this babble, as an extra precaution the transcripts of the recordings used were searched for the experimental target-words and none of these words occurred in the sections used.

The spectrally matched speech-shaped stochastic noise was created by Praat script (Boersma & Weenink, 2001), which measured the long-term average spectrum and the intensity profile of the multi-talker babble and applied this to stochastic noise of equal duration. The sounds in the two masker conditions thus had equivalent average spectra and identical intensity profiles.

Words Two lists of seven Arabic words (14 words altogether) were created as the target words that participants would try to identify in the noise. This is one word for each of the seven trials in each masker condition. The words in these lists were selected to use similar sounds (roughly equal numbers of nasals, sibilants and stops between the lists). This phonological balancing of the lists was an extra precaution to minimize variability in how easy these words were to hallucinate within each run of the experiment. The words were all chosen to be concrete, easily imageable bisyallbic nouns. For a given participant, each masker condition was assigned to a particular list, but the assignment of list to masker was switched for half of the participants, so the word-to-masker assignment was counterbalanced over the full run of the experiment. The lists of words are given in Table 1.

Table 1. Word lists for the two conditions. The Arabic is presented in Romanized form

List one		List two	
Arabic	English gloss	Arabic	English gloss
tuffah	'apple'	*katef*	'shoulder'
maktab	'desk'	*minshaar*	'saw'
shooka	'fork'	*furshaah*	'brush'
korsi	'chair'	*kitab*	'book'
mismaar	'nail'	*sikkin*	'knife'
shafra	'razor'	*sariir*	'bed'
bedla	'suit'	*dulab*	'cupboard'

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

2.2. Results

In debriefing, participants were asked if they had any ideas about the purpose of the experiment and whether they had realized that there were no recordings of the relevant words being played under the masker. No participant reported realizing that they were hallucinating the words that they reported hearing (and many expressed significant surprise when informed of this).

The dependent measure was the number of times a participant heard the target word in a given trial (AVH per trial) as determined by the number of spacebar hits per trial.

All statistical analyses were conducted with the **R** statistical software package (R Core Team, 2014). A generalized linear mixed model with a negative binomial distribution was fit using the *glmer* function in the lme4 package (Bates et al., 2015). The fixed effect was **Masker** type and the random effects structure specified random intercepts for **Participant** and random slopes for **Masker** and **Trial** number. This showed a significant effect of **Masker** type ($p < 0.0001$), with more AVH induced by *babble* than by *stochastic noise*. These results are shown (along with the results for **Experiment Two**) in Figure 3.

The analysis of the effect of **Masker** is taken up further in the combined results section below, which includes the results for **Experiment Two**.

All data collected (for both experiments) are available at the Open Science Framework: https://osf.io/hbkqy/.

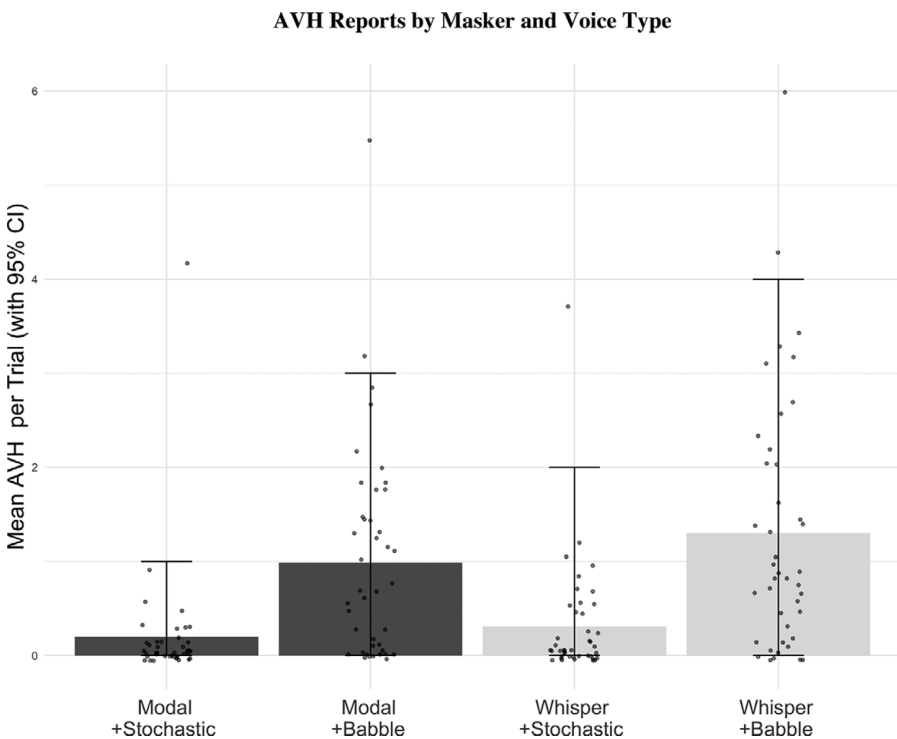


Figure 3. Results of experiments one and two combined. 95% Confidence Intervals are shown; individual data-points are also shown, jittered for visibility.

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

2.3. Discussion

These results show that hallucinations of speech are quite easy to induce (in line with previous work on the White Christmas effect) and that multi-talker babble is a more effective inducer of AVH than spectrally matched speech-shaped stochastic noise. Further discussion of the effect of **Masker** type is taken up in the general discussion.

3. Experiment two

Experiment Two replicates the masking effect explored in **Experiment One**, but with a change in the expected 'signal' side – testing whether an expectation of more easily masked speech (whisper) will show the same masking relations. **Experiment Two** also examines, by comparing results against **Experiment One**, whether the expectation of more-easily masked speech will lead to a higher rate of AVH. Comparison between expectation of whisper versus modal voice was done 'between-subjects' because it was felt that participants would find it difficult to switch their expectation of modal versus whispered speech – they would likely forget which kind of expectation they should have in a given block, thus muddling conditions. To avoid this problem, each participant was given only one type of speech to expect: modal voice (**Experiment One**) or whisper (**Experiment Two**).

The primary contributor to whisper's greater susceptibility to masking is the fact that whispered speech is simply inherently less intense than modal speech (Legou et al., 2010; Schwartz, 1970; Traunmüller & Eriksson, 2000). This results in whispered speech showing lower rates of intelligibility (Tartter, 1991).

When it comes to masking relationships, whisper is not just more easily masked because it is quieter but also because it lacks harmonic structure. Whispered speech is generated by creating a constriction in the larynx, likely involving the false vocal folds, but with suppression of vocal fold vibration (Hardcastle et al., 2012; Tsunoda et al., 1996). This laryngeal constriction creates turbulent airflow which excites resonances in the vocal tract above, but as there is no vocal fold vibration, there is no periodicity in whisper. In modal speech, the periodic source means that the component frequencies of the speech sound are integer multiples of the fundamental frequency (termed *harmonicity*); whisper, however, lacks this harmonic relation and so is *inharmonic* (Popham et al., 2018).

Harmonicity plays a central role in allowing perceivers to separate a target sound from background (Bregman, 1990; Nittrouer & Tarr, 2011): 'a common harmonic series helps to group together the different formant regions that make up a voiced sound such as a vowel' (Darwin, 2008, p.1016). This explains why periodic sounds are more resistant to masking from stochastic noise (Alm et al., 2009; Popham et al., 2018). This is presumably part of the reason for the lower rates of AVH for the speech-shaped stochastic noise masker in comparison to the multi-talker babble masker in **Experiment One**.

Given that whisper lacks harmonicity it should not show the same resistance to masking from stochastic noise that was shown in **Experiment One** for modal voice. We thus predicted that there should be an interaction of whisper with masking sound, whereby the difference in AVH between babble and spectrally matched speech-shaped stochastic noise shown in **Experiment One** should be reduced in **Experiment Two**.

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

A connection between whispered speech and AVH is highlighted by the fact that the voice hallucinations experienced by non-clinical populations typically sound to the hearer 'quiet (or like a whisper)' (Toh et al., 2020). This, of course, fits well with the signal-detection account of AVH, though, as discussed above, there may be several mechanisms by which AVH are caused.

3.1. Methods

The design of **Experiment Two** was identical to **Experiment One** except that participants were told that the words they were instructed to listen for would be whispered.

3.1.1. Participants

There were 42 new participants recruited at United Arab Emirates and Qatar University (none of the participants were the same as for **Experiment One**). All participants were native Arabic-speaking and female. All participants reported no known hearing problems. The average age was 20.34 years (SD = 3.38). Participants were paid or given course credit for their participation.

3.2. Results

In debriefing participants were asked if they had any ideas about the purpose of the experiment and whether they had realized that there were no recordings of the relevant words being played under the masker. No participant reported realizing that they were hallucinating the words that they reported hearing (and again, many reported surprise that there were no actual recordings of words).

As with **Experiment One**, a generalized linear mixed model with a negative binomial distribution was fit. The dependent measure was the average number of AVH per trial (as measured by the number of times each participant hit the spacebar per trial). The fixed effect was **Masker** type and the random effects structure specified random intercepts for **Participant** and random slopes for **Masker** type and **Trial** number. This showed a significant effect of **Masker** type ($p < 0.0001$), with more AVH induced by *babble* than by *stochastic noise*.

The effect of **Masker** type and **Voice expectation** type are reported in the combined results section below and a graph of these results (along with those from **Experiment One**) is shown in Figure 3.

3.3. Combined results of experiments one and two

To explore the effect of **Voice expectation** type, **Masker** type and their interaction, a generalized linear mixed model with a negative binomial distribution over the combined data for **Experiment One** and **Two** was fit. The dependent measure was the average number of AVH per trial. The fixed effects were **Masker** type and **Voice expectation** type and their interaction. The random effects specified random intercepts for **Participant** and random slopes for **Masker** type and **Trial** number. The results of this analysis are shown in Table 2. This found an effect of **Masker** type ($p < 0.0001$), and a slight non-significant trend for **Voice expectation** type ($p = 0.213$),

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

Table 2. Summary of GLM results

	Estimate	Std. Error	z	Pr(>/z/)
Intercept	−0.3889	0.2375	−1.637	0.102
Voice type	−0.4089	0.3282	−1.246	0.213
Masker type	−2.2207	0.3674	−6.044	<0.0001
Voice type × Masker type	−0.4470	0.4511	−0.991	0.322

with more AVH being reported for whisper in comparison to modal voice. Contrary to prediction, there was little evidence of an interaction of **Masker** type with **Voice expectation** type (p = 0.322). The results of both experiments are shown in Figure 3.

3.4. Discussion

These results replicate the effect of **Masker** type on the White Christmas effect found in **Experiment One** by demonstrating the same effect with an expectation of whispered speech. That AVH would be easier to induce when the expectation was of hearing whispered speech makes sense from a signal-detection perspective as whispered speech is easier to mask. However, while there was a slight trend toward a stronger effect for the whispered voice in **Experiment Two** in comparison to the modal voice in **Experiment One**, this did not reach significance. Furthermore, the predicted interaction of **Voice expectation** type (modal versus whispered) with **Masker** type (multi-talker babble versus spectrally matched speech-shaped stochastic noise) was not found.

4. General discussion

These results show that speech hallucinations are influenced by the type of masking sound and perhaps by phonetic expectation. With multi-talker babble masking, the rates of AVH are, as predicted, higher. Similarly, when the participant is told to expect whisper, AVH rates trend higher (non-significantly in this data). Multi-talker babble is a more effective masker than stochastic noise, and whisper is more easily masked than modal voice, and so these results support the view that the masking potential rule applies to the White Christmas effect. When the masker is sufficient to mask an expected word, listeners are more likely to hallucinate the expected word. These results mirror findings on phoneme restoration, which also follows the masking potential rule (e.g., Bashford et al., 1992; Bashford & Warren, 1987; Warren et al., 1972).

A consistent issue with the White Christmas effect is that participants may feel pressured, no matter how clear the instructions are, to indicate they heard the target word, even when they did not. In the experiments reported here, this pressure may (as with other experiments on the White Christmas effect) inflate the total number of 'hits'. However, it does not plausibly explain the rate of the effect *differing* across conditions.

Given the greater masking of voiced speech (in comparison to whisper) by babble than by stochastic noise, we predicted that there should be an interaction between **Masker** type (multi-talker babble versus spectrally-matched speech-shaped stochastic noise) and **Voice expectation** type (modal versus whisper). Specifically we

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

predicted that there should be a smaller difference between masking conditions for whisper than for modal speech. This interaction was not significant.

There are differences between modal and whispered speech other than periodicity. For example, whispered consonants are typically longer (Schwartz, 1972; Tartter, 1989; Xu et al., 2022), formants of vowels tend to be higher in frequency (Kallail & Emanuel, 1984a, b) and the vowel space for whispered vowels may shrink in comparison to modal speech (Houle & Levi, 2020). Of particular importance to masking, the spectral tilt of whisper emphasizes the higher frequency end of the spectrum and the intensity of vowels reduces disproportionately to consonants in whispered speech, altering the consonant-vowel intensity ratio, which plays a role in intelligibility under masking (Freyman et al., 1991, 2012). Thus, perhaps differences between modal and whispered speech other than voicing explain the lack of significant interaction between **Voice expectation** and **Masker** type in this data set.

Another possibility is that the likelihood of AVH is not perfectly aligned with the 'masking potential rule' – perhaps the auditory system's top-down expectations of whether a masker would mask a particular speech sound are not entirely accurate. If this is the case, then probing the ability of various qualities of masker to induce illusions such as the White Christmas effect or phoneme restoration would serve as a useful way to probe differences between the actual masking potential of types of masker and the auditory system's predictions of masking from these various types of sound.

One way in which the 'masking potential rule' may be insufficient to explain the White Christmas Effect and possibly other continuation effects such as phoneme restoration, is that it relies on *energetic* masking. Energetic masking refers to the decrease in audibility of a sound when there is a competing sound in the same frequency band (Fletcher, 1940; Pollack, 1975; Srinivasan & Wang, 2008). In contrast, *informational* masking is the "perceptual degradation caused by the listener's inability to segregate a target signal from interference" (Srinivasan & Wang, 2008, 3213). This form of masking may also play a role in continuity illusions. There is no actual signal to be masked (either energetically or informationally) in the White Christmas or phoneme restoration effects, but the 'masking potential rule' assumes that the perceptual system can determine whether a particular masker *would* have energetically masked a (non-existent) sound. It is possible that informational masking is also taken into account by the perceptual system – perhaps the system estimates how easy it would be to segregate a target sound against a given background and so would consider it more likely, via biasing in a signal-detection sense, that a sound occurred in a high information-masking context.

The concept of informational masking is complicated by the fact that the term is used variably by different authors and is something of a 'suitcase word' (Watson, 2005), lumping together all causes of masking that cannot be explained by mechanisms in the auditory periphery (which are instead considered to cause *energetic masking*). The list of mechanisms potentially contributing to informational masking includes, among others, similarity between masker and signal, stimulus/masker uncertainty and distraction. These factors are not mutually exclusive and may interact (as when stimulus uncertainty causes distraction). When multi-talker babble with a small number of speakers is used as a masker, informational masking is high because many of these mechanisms are at play – there is essentially a *cocktail party* problem (Cherry, 1953) of segregating the target from the background. As the number of speakers in the babble increases, individual distractor voices are harder

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

to pick out and so informational masking is reduced and energetic masking dominates (Leibold & Buss, 2013). Some researchers would argue that when speech is no longer intelligible in multi-talker babble, the masking should no longer be labeled *informational masking*. While energetic masking does dominate informational masking when there are a high number of speakers in multi-talker babble, the problem of similarity between masker and target remains (to a degree) as can the problem of stimulus uncertainty. These would not normally be considered energetic masking. Whether this should be labeled information masking is a debate to which these experiments do not make a contribution. If information masking (in this restricted target-masker similarity sense) does play a role in the White Christmas effect, then that would predict that AVH should be more common for expecting whispered speech in the stochastic-noise condition (in comparison to the babble condition); and more common when expecting modal speech in the babble condition (in comparison to the stochastic-noise condition). Both of these situations involve more similarity between (expected) target and masking sound – greater similarity between target and masker should lead to more informational masking (Brungart et al., 2001; Kidd et al., 2008). An experiment conducted by Buss et al. (2022) followed a similar logic, arguing that whispered speech should be easier to detect against a voiced masker and voiced speech should be easier to detect against a whispered masker. Their argument was based on the assumption that greater dissimilarity between target and masker should reduce informational masking. If this form of informational masking played a role in the experiments reported above, it would push the pattern of AVH in the opposite direction to the interaction that we predicted and so may explain the absence of the predicted interaction between **Voice expectation** and **Masker** type.

The results of these experiments are consistent with a signal-detection bias interpretation of AVH – keeping in mind that this signal-detection bias is provided by top-down expectations which could come from a memory or from mental imagery.

A link between mental imagery and hallucination was suggested by Ulric Neisser (1976, 1978), who argued that perceptual expectations (which he called schemata) are normally used to shape perception but can be triggered, in the absence of external stimulation, to constitute imagery. Recent experiments have demonstrated an overlap between imagery and hallucination, showing that there is a tendency to hear both imagined speech and AVH induced through white noise as occurring in the right ear. Prete et al. (2016, 2018) in line with the well-established 'Right Ear Advantage' (REA) for the perception of external speech (Broadbent, 1954; Hugdahl, 2011). This REA for imagined speech sounds was recently replicated in an rTMS study (Prete et al., 2024). However, a study investigating REA among voice-hearers with schizophrenia did not find such lateralization for either real or imagined speech (Altamura et al., 2020). The 2018 study by Prete et al. showed an interesting interaction with intensity, as participants were more likely to experience an REA for syllables presented at a lower intensity in comparison to a masking noise (though this was not a White Christmas effect, as there was a syllable spoken in the noise).

Variations of the idea of a common source for imagery and hallucination have been put forward by several authors, (e.g., Clark, 2013; Grush, 2004). In a recent variant of this idea, sensory predictions generated by the motor system (known as *forward models*) have been argued to play a role in both perception and mental imagery. Forward models have a primary use in motor control and in preventing

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

confusion between self-caused and externally caused sensations (Aliu et al., 2009), but it has been argued that forward models can be run 'offline' (as in Neisser's *schemata* discussed above) to generate speech imagery which, in turn, has been shown to influence the perception of external speech (Scott, 2016; Scott et al., 2013). Malfunction of this forward model system has been proposed as the origin of voice hearing in schizophrenia (Frith, 1992). This shaping of ambiguous speech by expectations generated by the motor system can even be caused by the static position of speech articulators (Yeung & Scott, 2021). Similarly, Sato et al. (2006) have found a link between articulation and the verbal transformation effect, which, as discussed above, is very similar to the White Christmas effect. A useful summary and synthesis of the role of prediction mechanisms in perception, imagery and hallucination is presented in Clark (2013).

The role of predictive mechanisms in perception, imagery and hallucination has more recently been cast in Bayesian terms (e.g., Cassidy et al., 2018; Fletcher & Frith, 2009; Powers et al., 2017). The brain is argued to use a Bayesian-like weighting of incoming sensory stimulation and prior probabilities to achieve the most plausible percept despite noisy and underdetermined sensory input. Thus, perception is dependent on both top-down (expectation) and bottom-up (sensory stimulation) factors. Under this framework, we would predict that both the details of what is expected and the details of the incoming sensory stimulus to play a role in the likelihood of a hallucinatory experience, as supported by the current experiments.

An alternative to a masking explanation of why babble induces higher rates of AVH is that babble more strongly engages the speech centres of the brain and thus makes the experience of hearing speech more likely. This would be similar to the proposed attentional explanation of the REA, by which simply engaging the left hemisphere by drawing a person's attention to the right side of space is sufficient to boost language processing (Kinsbourne, 1970).

4.1. Limitations and future research

The results of these experiments suggest the possible involvement of informational masking, but were not specifically designed to address this involvement. Future research will use variations in babble masking, with a smaller number of talkers in the babble, to explore this possibility.

The current experiments did not take a baseline measure of participants' proneness to hallucination or require disclosure of any mental health conditions. This was for reasons of both privacy and experiment brevity. It would be interesting to explore, in future research, whether the masking relations demonstrated in the experiments above are influenced by participant proneness to hallucination. As part of this, there was no objective measure of participant hearing threshold. It is possible that proneness to hallucination is influenced by overall hearing acuity and so future research will include such a measure.

It is likely that semantic and word-frequency effects play a role in the probability of hearing a particular word in the White Christmas effect, but the experiments reported here concentrate on how the *auditory* aspects of potential masking modulate verbal hallucinations. While the experiments reported above do not directly address any issues specific to the Arabic language, future research is planned which will use this White Christmas methodology to explore the special morphology of Arabic –

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

examining whether it is easier to induce this effect for words that fit the standard *triliteral consonant* morphological structure of Arabic in comparison to equally common words which do not have this structure. Through collaborations with other institutions, we also hope to address the sex-balance issue: We were not able to collect a sex-balanced sample because the departments from which participants were recruited only have female students. This limits the generalizability of our results, something we hope to address in future research.

5. Conclusions

These experiments show that masking relations modulate the White Christmas effect, in line with predictions based on the phoneme restoration literature. Masking from multi-talker babble led to a higher rate of verbal hallucinations than masking from spectrally matched speech-shaped stochastic noise. Similarly, there was a trend toward a higher rate of verbal hallucinations when participants expected to hear a more easily masked sound (whisper) than for 'normal' modal speech. Contrary to prediction, no significant interaction was found between these factors.

Acknowledgments. Participants were run by research assistants Maryam Moustafa Gadalla Aref, Fatima Boush and Rofida Babiker Hamid Ibrahim. Preliminary results from a subset of this data were presented at the conference of the Canadian Acoustics Association (Scott, 2022).

Funding statement. This research was funded by a Qatar University grant QUUG-CAS-ELL-17_18–7, and publication was made possible through the generous sponsorship of the publication fee by United Arab Emirates University.

References

Alganami, F., Varese, F., Wagstaff, G. F., & Bentall, R. P. (2017). Suggestibility and signal detection performance in hallucination-prone students. *Cognitive Neuropsychiatry*, 22(2), 159–174.

Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience*, 21(4), 791–802.

Alm, M., Behne, D. M., Wang, Y., & Eg, R. (2009). Audio-visual identification of place of articulation and voicing in white and babble noise. *Journal of the Acoustical Society of America*, 126(1), 377–387.

Altamura, M., Prete, G., Elia, A., Angelini, E., Padalino, F. A., Bellomo, A., Tommasi, L., & Fairfield, B. (2020). Do patients with hallucinations imagine speech right? *Neuropsychologia*, 146, 107567.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.) American Psychiatric Association.

Barber, T. X., & Calverley, D. S. (1964). Empirical evidence for a theory of "hypnotic" behavior: Effects of pretest instructions on response to primary suggestions. *The Psychological Record*, 14(4), 457–467.

Bashford, J. A., Riener, K. R., & Warren, R. M. (1992). Increasing the intelligibility of speech through multiple phonemic restorations. *Perception & Psychophysics*, 51(3), 211–217.

Bashford, J. A., & Warren, R. M. (1987). Multiple phonemic restorations follow the rules for auditory induction. *Perception & Psychophysics*, 42(2), 114–121.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Bentall, R. P., Baker, G. A., & Havers, S. (1991). Reality monitoring and psychotic hallucinations. *British Journal of Clinical Psychology*, 30(3), 213–222.

Bentall, R. P., & Slade, P. D. (1985). Reality testing and auditory hallucinations: A signal detection analysis. *British Journal of Clinical Psychology*, 24(3), 159–169.

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10), 341–345.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press.

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 47(3), 191–196.

Brookwell, M. L., Bentall, R. P., & Varese, F. (2013). Externalizing biases and hallucinations in source-monitoring, self-monitoring and signal detection studies: A meta-analytic review. *Psychological Medicine*, 43(12), 2465–2475.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, 110(5), 2527–2538.

Buss, E., Miller, M. K., & Leibold, L. J. (2022). Maturation of speech-in-speech recognition for whispered and voiced speech. *Journal of Speech, Language, and Hearing Research*, 65(8), 3117–3128.

Cassidy, C. M., Balsam, P. D., Weinstein, J. J., Rosengard, R. J., Slifstein, M., Daw, N. D., Abi-Dargham, A., & Horga, G. (2018). A perceptual inference mechanism for hallucinations linked to striatal dopamine. *Current Biology*, 28(4), 503–514.e4.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 25, 975–979.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204.

Crowe, S., Barot, J., Caldow, S., D'Aspromonte, J., Dell'Orso, J., Di Clemente, A., Hanson, K., Kellett, M., Makhlota, S., McIvor, B., McKenzie, L., Norman, R., Thiru, A., Twyerould, M., & Sapega, S. (2011). The effect of caffeine and stress on auditory hallucinations in a non-clinical sample. *Personality and Individual Differences*, 50(5), 626–630.

Darwin, C. J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363(1493), 1011–1021.

de Cheveigné, A., McAdams, S., Laroche, J., & Rosenberg, M. (1995). Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement. *Journal of the Acoustical Society of America*, 97(6), 3736–3748.

Deutsch, D. (2003). *Phantom words and other curiosities [album]*. Philomel Records.

Deutsch, D. (2019). *Musical illusions and phantom words: How music and speech unlock mysteries of the brain*. Oxford University Press.

Drouin, M., Kaiser, D. H., & Miller, D. A. (2012). Phantom vibrations among undergraduates: Prevalence and associated psychological characteristics. *Computers in Human Behavior*, 28(4), 1490–1496.

Feelgood, S. R., & Rantzen, A. J. (1994). Auditory and visual hallucinations in university students. *Personality and Individual Differences*, 17(2), 293–296.

Fernyhough, C., Bland, K., Meins, E., & Coltheart, M. (2007). Imaginary companions and young children's responses to ambiguous auditory stimuli: Implications for typical and atypical development. *Journal of Child Psychology and Psychiatry*, 48(11), 1094–1101.

Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, 12(1), 47–65.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10(1), 48–58.

Freyman, R. L., Griffin, A. M., & Oxenham, A. J. (2012). Intelligibility of whispered speech in stationary and modulated noise maskers. *The Journal of the Acoustical Society of America*, 132(4), 2514–2523.

Freyman, R. L., Nerbonne, G. P., & Cote, H. A. (1991). Effect of consonant-vowel ratio modification on amplitude envelope cues for consonant recognition. *Journal of Speech, Language, and Hearing Research*, 34(2), 415–426.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Lawrence Erlbaum Associates.

Frost, R., Repp, B. H., & Katz, L. (1988). Can speech perception be influenced by simultaneous presentation of print? *Journal of Memory and Language*, 27(6), 741–755.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27(3), 377–442.

Hardcastle, W. J., Laver, J., & Gibbon, F. E. (2012). *The handbook of phonetic sciences*. John Wiley & Sons.

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

Hartley, S., Bucci, S., & Morrison, A. P. (2017). Rumination and psychosis: An experimental, analogue study of the role of perseverative thought processes in voice-hearing. *Psychosis*, 9(2), 184–186.

Hoffman, D. D. (2012). The construction of visual reality. In J. D. Blom & I. E. Sommer (Eds.), *Hallucinations: Research and practice* (pp. 7–15). Springer.

Hoskin, R., Hunter, M. D., & Woodruff, P. W. R. (2014). The effect of psychological stress and expectation on auditory perception: A signal detection analysis. *British Journal of Psychology*, 105(4), 524–546.

Houle, N., & Levi, S. V. (2020). Acoustic differences between voiced and whispered speech in gender diverse speakers. *The Journal of the Acoustical Society of America*, 148(6), 4002–4013.

Hugdahl, K. (2009). "Hearing voices": Auditory hallucinations as failure of top-down control of bottom-up perceptual processes. *Scandinavian Journal of Psychology*, 50(6), 553–560.

Hugdahl, K. (2011). Fifty years of dichotic listening research – Still going and going and. *Brain and Cognition*, 76(2), 211–213.

Kallail, K. J., & Emanuel, F. W. (1984a). An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects. *Journal of Phonetics*, 12(2), 175–186.

Kallail, K. J., & Emanuel, F. W. (1984b). Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects. *Journal of Speech, Language, and Hearing Research*, 27(2), 245–251.

Kashino, M. (2006). Phonemic restoration: The brain creates missing speech sounds. *Acoustical Science and Technology*, 27(6), 318–321.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In R. R. Fay, A. N. Popper, W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (Vol. 29, pp. 143–189). Springer.

Kim, J., & Davis, C. (2007). Restoration effects in auditory and visual speech. In J. Vroomen, M. Swerts, & E. Krahmer (Eds.), *Proceedings of the auditory–visual speech processing conference*. (pp. L6) Hilvarenbeek.

Kinsbourne, M. (1970). The cerebral basis of lateral asymmetries in attention. *Acta Psychologica*, 33, 193–201.

Legou, T., Ghio, A., Ch, A., & Giovanni, A. (2010). Etudes expérimentales préliminaires de la voix chuchotée: Pression sous-glottique et étude posturale. *Revue de Laryngologie Otologie Rhinologie*, 131(1), 1–4.

Leibold, L. J., & Buss, E. (2013). Children's identification of consonants in a speech-shaped noise or a two-talker masker. *Journal of Speech, Language, and Hearing Research*, 56(4), 1144–1155.

Merckelbach, H., & van de Ven, V. (2001). Another white Christmas: Fantasy proneness and reports of 'hallucinatory experiences' in undergraduate students. *Journal of Behavior Therapy and Experimental Psychiatry*, 32, 137–144.

Moseley, P., Smailes, D., Ellison, A., & Fernyhough, C. (2016). The effect of auditory verbal imagery on signal detection in hallucination-prone individuals. *Cognition*, 146, 206–216.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. Henry Holt & Co.

Neisser, U. (1978). Anticipations, images, and introspection. *Cognition*, 6(2), 169–174.

Nittrouer, S., & Tarr, E. (2011). Coherence masking protection for speech in children and adults. *Attention, Perception, & Psychophysics*, 73(8), 2606–2623.

Pollack, I. (1975). Auditory informational masking. *Journal of the Acoustical Society of America*, 57(S1), S5.

Popham, S., Boebinger, D., Ellis, D. P. W., Kawahara, H., & McDermott, J. H. (2018). Inharmonic speech reveals the role of harmonicity in the cocktail party problem. *Nature Communications*, 9(1), 2122.

Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science*, 357(6351), 596–600.

Prete, G., D'Anselmo, A., Brancucci, A., & Tommasi, L. (2018). Evidence of a right ear advantage in the absence of auditory targets. *Scientific Reports*, 8(1), 15569.

Prete, G., Marzoli, D., Brancucci, A., & Tommasi, L. (2016). Hearing it right: Evidence of hemispheric lateralization in auditory imagery. *Hearing Research*, 332, 80–86.

Prete, G., Rollo, B., Palumbo, R., Ceccato, I., Mammarella, N., Di Domenico, A., Capotosto, P., & Tommasi, L. (2024). Investigating the effect of rTMS over the temporoparietal cortex on the right ear advantage for perceived and imagined voices. *Scientific Reports*, 14(1), 24930.

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

Rassin, E., & van der Heiden, S. (2006). Swearing voices: An experimental investigation of the suppression of hostile hallucinations. *Behavioural and Cognitive Psychotherapy*, 35(3), 355–360.

Reisberg, D. (1989). Enacted" auditory images are ambiguous; "pure" auditory images are not. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 41A(3), 619–641.

Samuel, A. G. (1996). Phoneme restoration. *Language and Cognitive Processes*, 11(6), 647–654.

Sato, M., Schwartz, J.-L., Abry, C., Cathiard, M.-A., & Loevenbruck, H. (2006). Multistable syllables as enacted percepts: A source of an asymmetric bias in the verbal transformation effect. *Perception & Psychophysics*, 68(3), 458–474.

Schwartz, M. F. (1970). Power spectral density measurements of oral and whispered speech. *Journal of Speech and Hearing Research*, 13(2), 445–446.

Schwartz, M. F. (1972). Bilabial closure durations for /p/, /b/, and /m/ in voiced and whispered vowel environments. *The Journal of the Acoustical Society of America*, 51(6B), 2025–2029.

Scott, M. (2016). Speech imagery recalibrates speech-perception boundaries. *Attention, Perception & Psychophysics*, 78(5), 1496–1511.

Scott, M. (2022). A whispered Christmas: Phonetic expectations and type of masking-noise influence auditory verbal hallucinations. *Journal of the Canadian Acoustical Association*, 50(3), 88–89.

Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *Journal of the Acoustical Society of America Express Letters*, 133(4), 286–293.

Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *Journal of the Acoustical Society of America*, 118(5), 2775–2778.

Slade, P. D., & Bentall, R. P. (1988). *Sensory deception: A scientific analysis of hallucination*. Johns Hopkins University Press.

Srinivasan, S., & Wang, D. (2005). A schema-based model for phonemic restoration. *Speech Communication*, 45(1), 63–87.

Srinivasan, S., & Wang, D. (2008). A model for multitalker speech perception. *The Journal of the Acoustical Society of America*, 124(5), 3213–3224.

Steinmetzger, K., & Rosen, S. (2015). The role of periodicity in perceiving speech in quiet and in background noise. *Journal of the Acoustical Society of America*, 138(6), 3586–3599.

Tartter, V. C. (1989). What's in a whisper? *Journal of the Acoustical Society of America*, 86(5), 1678–1683.

Tartter, V. C. (1991). Identifiability of vowels and speakers from whispered syllables. *Perception & Psychophysics*, 49(4), 365–372.

Toh, W. L., Moseley, P., & Fernyhough, C. (2022). Hearing voices as a feature of typical and psychopathological experience. *Nature Reviews Psychology*, 1(2), 72–86.

Toh, W. L., Thomas, N., Robertson, M., & Rossell, S. L. (2020). Characteristics of non-clinical hallucinations: A mixed-methods analysis of auditory, visual, tactile and olfactory hallucinations in a primary voice-hearing cohort. *Psychiatry Research*, 289, 112987.

Traunmüller, H., & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America*, 107, 3438–3451.

Trout, J., & Poser, W. J. (1990). Auditory and visual influences on phonemic restoration. *Language and Speech*, 33(2), 121–135.

Tsunoda, K., Niimi, S., & Hirose, H. (1996). The roles of the posterior cricoarytenoid and thryopharyngeus muscles in whispered speech. *Folia Phoniatrica et Logopaedica*, 46, 139–151.

van Slobbe-Maijer, K. (2019). *Auditory hallucinations in youth: Occurrence, clinical significance and intervention strategies*. University of Groningen.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167(3917), 392–393.

Warren, R. M., & Gregory, R. L. (1958). An auditory analogue of the visual reversible figure. *The American Journal of Psychology*, 71(3), 612–613.

Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176(4039), 1149–1151.

Watson, C. S. (2005). Some comments on informational masking. *Acta Acustica United with Acustica*, 91, 502–512.

Xu, M., Shao, J., Ding, H., and Wang, L. (2022). Acoustic-perceptual correlates of whispered mandarin consonants. In *2022 13th international symposium on Chinese spoken language processing (ISCSLP)* (pp. 195–199). IEEE.

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press

Yeung, H. H., & Scott, M. (2021). Postural control of the vocal tract affects auditory speech perception. *Journal of Experimental Psychology: General*, 150(5), 983–995.

Young, H. F., Bentall, R. P., Slade, P. D., & Dewey, M. E. (1987). The role of brief instructions and suggestibility in the elicitation of auditory and visual hallucinations in normal and psychiatric subjects. *The Journal of Nervous and Mental Disease*, 175(1), 41–48.

Cite this article: Scott, M., & Leung, T. T.-C. (2025). I'm whispering a white Christmas: masking relations in hallucinatory speech, *Language and Cognition*, 17, e71, 1–20. https://doi.org/10.1017/langcog.2025.10030

https://doi.org/10.1017/langcog.2025.10030 Published online by Cambridge University Press