Talkers produce more pronounced amplitude modulations when speaking in noise

Hans Rutger Bosker, and Martin Cooke

Citation: The Journal of the Acoustical Society of America **143**, EL121 (2018); doi: 10.1121/1.5024404
View online: https://doi.org/10.1121/1.5024404
View Table of Contents: http://asa.scitation.org/toc/jas/143/2
Published by the Acoustical Society of America

Hans Rutger Bosker and Martin Cooke: JASA Express Letters https://doi.org/10.1121/1.5024404 Published Online 15 February 2018

Talkers produce more pronounced amplitude modulations when speaking in noise

Hans Rutger Bosker[a)]
Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH, Nijmegen, The Netherlands
HansRutger.Bosker@mpi.nl

Martin Cooke[b)]
Language and Speech Laboratory, Universidad del País Vasco, Vitoria, 01006, Spain
m.cooke@ikerbasque.org

Abstract: Speakers adjust their voice when talking in noise (known as Lombard speech), facilitating speech comprehension. Recent neurobiological models of speech perception emphasize the role of amplitude modulations in speech-in-noise comprehension, helping neural oscillators to "track" the attended speech. This study tested whether talkers produce more pronounced amplitude modulations in noise. Across four different corpora, modulation spectra showed greater power in amplitude modulations below 4 Hz in Lombard speech compared to matching plain speech. This suggests that noise-induced speech contains more pronounced amplitude modulations, potentially helping the listening brain to entrain to the attended talker, aiding comprehension.

[RS]
Date Received: November 14, 2017 **Date Accepted:** January 31, 2018

1. Introduction

Speakers typically adjust their voice when talking in noisy conditions. Speech produced in noise generally exhibits increased intensity, slower speech rate, raised F0, and flatter spectral tilt (for an overview, see Cooke *et al.*, 2014a). These and other modifications result in what is collectively known as Lombard speech (Lombard, 1911). The scientific importance of this form of speech stems in large part from the discovery by Dreher and O'Neill (1957) that, after discounting intensity increases, Lombard speech is more intelligible than unmodified speech when presented in noise, a finding that has been confirmed in a number of subsequent studies (e.g., Pittman and Wiley, 2001; Summers *et al.*, 1988). However, the basis for intelligibility gains is not fully understood. One aspect of noise-induced speech that has received little attention concerns how talkers adjust the temporal modulations of their speech when conversing in noise. The present study examines the temporal modulations in Lombard speech and plain speech (speech produced in quiet) and demonstrates that amplitude modulations are enhanced in the temporal envelope of Lombard speech compared to matching plain speech.

Speech is an inherently rhythmic signal in that it contains strong amplitude modulations, particularly in the 1–15 Hz range (Ding *et al.*, 2017; Varnet *et al.*, 2017). These amplitude modulations, evident in the temporal envelope of speech, greatly contribute to speech intelligibility (Drullman *et al.*, 1994; Ghitza, 2012; Shannon *et al.*, 1995; Smith *et al.*, 2002). Speech with more pronounced amplitude modulations is more intelligible in noise (Houtgast and Steeneken, 1985; Koutsogiannaki and Stylianou, 2016; Steeneken and Houtgast, 1980). Also, speakers who are intrinsically more intelligible than others show more pronounced low-frequency modulations in the temporal envelope (Bradlow *et al.*, 1996). Modulations as low as 2 Hz have been shown to be essential for phoneme identification (Drullman *et al.*, 1994). In fact, removing amplitude modulations, occurring at a syllabic rate, from speech impairs its intelligibility to a large degree (Ghitza, 2012).

Recent neurobiological models of speech perception (Ghitza, 2011; Giraud and Poeppel, 2012; Peelle and Davis, 2012) propose that enhanced temporal modulations facilitate speech perception because speech-envelope information evokes marked "envelope-following" neural responses in the auditory cortex, known as *neural entrainment*.

[a)]Author to whom correspondence should be addressed. Also at: Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands.
[b)]Also at: Ikerbasque (Basque Science Foundation), Bilbao, Spain.

Endogenous neural oscillators in the delta (1–4 Hz) and theta range (4–8 Hz) are thought to phase-lock (entrain) to the amplitude fluctuations in the input signal (Doelling *et al.*, 2014). Thus, neuronal excitability is temporally aligned with the temporal structure of the attended spoken input, serving as a parsing mechanism for the initial neural coding of the speech signal (Arnal *et al.*, 2015; Bosker, 2017; Bosker and Kösem, 2017; Kösem *et al.*, 2017).

This neural entrainment to the temporal modulations in speech has been proposed to be one of the mechanisms by which listeners are capable of understanding speech in challenging listening conditions, such as in noise or with competing talkers. Brain oscillations during speech-in-noise perception preferentially track attended relative to ignored speech streams, using particularly the phase of low-frequency neural activity (1–8 Hz) (Ding and Simon, 2012; Kerlin *et al.*, 2010). The intelligibility of an attended speech stream in noise can be predicted from the extent to which cortical oscillators are aligned to the temporal envelope of the attended signal (Golumbic *et al.*, 2013; Golumbic *et al.*, 2012; Rimmele *et al.*, 2015). In fact, modulating listeners' neural activity with transcranial stimulation with speech-envelope-shaped currents has been argued to help speech-in-noise comprehension (Riecke *et al.*, 2018).

Considering these neurobiological models and the reported beneficial effects of enhanced amplitude modulations on perception, it may be hypothesized that speakers, in an attempt to aid speech intelligibility, would also naturally produce more enhanced amplitude modulations when talking in a noisy acoustic environment. This would allow greater opportunity for the listening brain to entrain to the enhanced temporal envelope. There is some evidence for larger within-syllable intensity changes (Garnier and Henrich, 2014) and greater overall RMS range (Folk and Schiel, 2011) in Lombard speech (relative to speech-in-quiet) but none of these studies examined the strength of amplitude modulations in the temporal envelope.

Krause and Braida (2004) investigated another kind of speech adjustment, namely, *clear speech*. In contrast to Lombard speech, clear speech is elicited in quiet environments by explicitly asking speakers to speak more clearly (e.g., by imagining speaking to a hearing-impaired person; Uchanski, 2008). Modulation spectra, showing the power of frequency components in the temporal envelope, revealed stronger amplitude modulations below 4 Hz in clear speech. This difference between clear and plain speech was most apparent in frequency bands around 500, 1000, and 2000 Hz. However, the effect was only observed for two talkers (T3 and T5; 50 trials each) and was induced through instructions rather than by physically presented adverse listening conditions (Krause and Braida, 2002).

Therefore, the present study examines the temporal modulations in Lombard speech and plain speech, adapting the method from Krause and Braida (2004). Using modulation spectra, the power of amplitude modulations in the temporal envelope of Lombard speech and plain speech is compared in three modulation frequency bands: the delta range (1–4 Hz), the theta range (4–8 Hz), and the alpha range (8–15 Hz).[1] In neurobiological studies, neural entrainment is particularly observed in the lower frequency range. Accordingly, we hypothesize that, across several speech corpora, Lombard speech will have more pronounced amplitude modulations compared to plain speech in the delta/theta range, as evidenced by greater power in the modulation spectrum.

2. Method

Four English speech corpora were analyzed (see Table 1), each including sentences produced in quiet and the same sentences produced in noisy elicitation conditions. The corpora varied on several dimensions: for instance, corpus 1 and 3 used "normal" sentences (i.e., meaningful everyday sentences), while corpus 2 used six-word matrix

Table 1. Characteristics of the four speech corpora [M = male; F = female; BMN = 9-talker babble-modulated ICRA noise from Dreschler *et al.* (2001); SSN = speech-shaped noise; SMN = speech-modulated noise].

	Talkers	Sentences	Noise	Source
Corpus 1	$N = 1$ (M)	"Normal"; $N = 25$	BMN; intense	Mayo *et al.* (2012)
Corpus 2	$N = 8$ (4 M/4 F)	Matrix; $N = 400$	SSN; 96 dB (L)	Lu and Cooke (2008)
Corpus 3 (Hurricane)	$N = 1$ (M)	"Normal"; $N = 720$	SMN; 84 dB (A)	Cooke *et al.* (2013)
Corpus 4 (MRT)	$N = 1$ (M)	Frame; $N = 300$	SMN; 84 dB (A)	Collected by Valentino-Botinhao (2013)

sentences (e.g., "lay green with A4 now" or "set white at B8 again") and corpus 4 used frame sentences [e.g., "Now we will say CV(C)C again"]. Corpora also varied in the noise conditions and loudness levels used to elicit Lombard speech; for instance, some used speech-shaped noise (i.e., noise with speech-like LTAS), others used noise modulated by a single talker or multiple talkers (e.g., ICRA noise from Dreschler *et al.*, 2001).

Before analysis, any leading and trailing silences around the sentences were manually removed. Two types of analysis were performed: a broadband analysis and a filterbank analysis. Both the broadband analysis and the filterbank analysis involved calculating the modulation spectrum of each sentence in each corpus using a method adapted from Krause and Braida (2004). This included normalizing the overall power of signals [root-mean-square (RMS)], matching the overall energy of plain and Lombard signals. Thus, any potential differences between plain and Lombard speech cannot be attributed to differences in overall energy.

For the broadband analysis, each sentence was filtered by a sixth-order Butterworth band-pass filter spanning the 250–4000 Hz range, followed by estimation of the envelope of the filter's output via the Hilbert transform. The envelope signal was then submitted to a fast Fourier transform, resulting in the modulation spectrum of that particular speech fragment. Finally, for statistical comparisons, the average power in three frequency bands was calculated: average power in the 1–4 Hz range (delta), the 4–8 Hz range (theta), and the 8–15 Hz range (alpha). This resulted in three different observations for each sentence, forming the dependent variables for the statistical analyses reported below.

The filterbank analysis was performed to further investigate whether any potential difference between the plain and Lombard speech could be attributed to particular frequency bands. The filterbank analysis was identical to the broadband analysis, except that the speech signal was filtered into five component signals, using a bank of fourth-order Butterworth filters with center frequencies of 500 Hz (bandwidth: 125 Hz), 1000 Hz (250 Hz), 2000 Hz (500 Hz), 4000 Hz (1000 Hz), and 8000 Hz (2000 Hz). The bandlimited output of this filterbank formed the input for the subsequent calculation of the modulation spectrum separately for each frequency band (using the procedure described above).

Since each sentence had its own unique duration, this procedure resulted in modulation spectra with unique frequency resolutions. However, in order to visualize the *average* rhythmicity in plain and Lombard speech across multiple sentences, identical frequency resolutions were required. This was achieved by repeating the procedure above with the envelope signal zero-padded to the next power of 2 higher than the length of the longest fragment of that particular corpus. Note that this zero-padding was only performed for visualization purposes; statistical analyses were performed on the data from the non-padded signals.

3. Results

Figure 1 shows the average modulation spectra from the broadband analysis (250–4000 Hz) for plain and Lombard speech, for each corpus.

The first thing to note is that the modulation spectra of Lombard speech resemble those of plain speech in overall shape (e.g., number of peaks and troughs), indicating that the temporal structure of matching Lombard and plain utterances is very much alike. Second, the peaks in the modulation spectra of Lombard speech occur at slightly lower frequencies (i.e., shifted leftward) than the peaks in the modulation spectra of plain speech. This observation is in line with the fact that Lombard speech typically has a slower speech rate than plain speech, shifting the temporal modulations down towards lower frequencies. In fact, the size of the shift observed in the modulation spectra is in

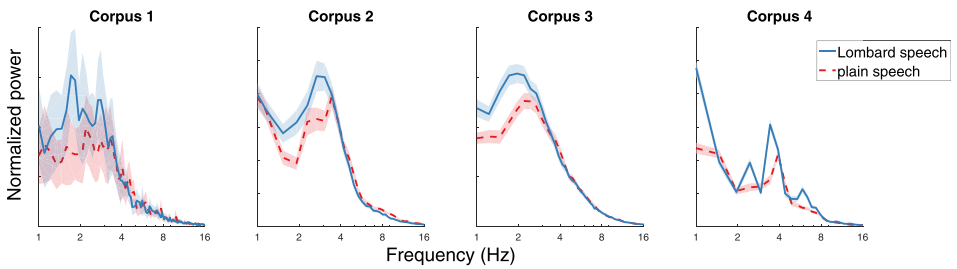

Fig. 1. (Color online) Average modulation spectra, calculated from the broadband analysis (250–4000 Hz), of Lombard speech (solid line) and matching plain speech (dashed line), for each corpus. Shaded areas indicate 95% CIs.

keeping with the average rate change in Lombard speech (e.g., 7.6% in corpus 2; Lu and Cooke, 2008).

Finally, there would seem to be consistently higher power in Lombard speech, across the four corpora, in the lower frequency range between 1–4 Hz (delta). This difference was statistically analyzed by means of linear mixed models (LMMs) (Baayen, 2008) as implemented in the lme4 library in R. Our three dependent variables, average power in delta (1–4 Hz), theta (4–8 Hz), and alpha range (8–15 Hz), were entered into separate LMMs with identical structure. Condition (categorical predictor, with the plain condition mapped onto the intercept), Corpus (categorical predictor, with corpus 1 mapped onto the intercept), and their interaction, were entered as predictors, with Talker entered as random factor with by-talker random slopes for Condition (Barr *et al.*, 2013). More complex models including a by-talker random slope for Condition failed to converge. Statistical significance was assessed at the 0.05 significance level by checking whether effects had absolute *t*-values exceeding 2 (Baayen, 2008).

Only the model of average power *in the delta range* (marginal $R^2 = 0.185$; conditional $R^2 = 0.408$) revealed a significant effect: Lombard speech had a significantly higher average power in the delta range compared to plain speech ($\beta = 0.008$, $SE = 0.002$, $t = 3.855$). The fact that no significant interactions between condition and corpus were observed suggests that the effect of Condition held equally across all four corpora. No significant effects or interactions were observed in the other two models.

In order to explore whether this difference in overall power between Lombard and plain speech in the delta range happened to be a by-product of the lower average speech rate in Lombard speech, a follow-up analysis was performed. This analysis involved artificially matching the (slower) speech rate of the Lombard sentences to the (faster) speech rate of the plain sentences. We adopted the global duration modifications described in Cooke *et al.* (2014b), involving linear compression of the Lombard utterances using PSOLA in PRAAT. The results of statistical analyses of the modulation spectra of plain and these duration-matched Lombard sentences mirrored the results reported above. We found no significant effects in the theta or alpha range, and only one significant effect of Condition in the delta range ($\beta = 0.011$, $SE = 0.005$, $t = 2.230$; model's marginal $R^2 = 0.127$; conditional $R^2 = 0.226$), corroborating that duration-matched Lombard utterances had higher overall power in the delta range than plain utterances across the four corpora.

The filterbank analysis was designed to further investigate which frequency bands drive the difference in temporal modulations between plain and Lombard speech. Figure 2 shows the average modulation spectra for plain and Lombard speech from the filterbank analysis (with center frequencies at 500–8000 Hz), averaging over the four different corpora.

Judging from Fig. 2, higher overall power in Lombard speech in the delta range would seem to be primarily driven by the 1000 and 2000 Hz bands. Differences between Lombard and plain speech in the delta range were statistically tested by means of another LMM. This LMM predicted average power in the delta range (1–4 Hz) with fixed effects of Condition (categorical predictor, with the plain condition mapped onto the intercept), Band (categorical predictor, with the 2000 Hz frequency band mapped onto the intercept), and their interaction. Talker was entered as random factor with by-talker random slopes for condition and band (marginal $R^2 = 0.686$; conditional $R^2 = 0.773$). More complex models including the predictor Corpus failed to converge.

A simple effect of Condition revealed significantly higher average power in the Lombard condition (relative to plain) in the delta range in the 2000 Hz frequency band (being mapped onto the intercept; $\beta = 0.003$, $SE = 0.0002$, $t = 14.850$). Interactions between Condition and Band led to two observations: first, the absence of an interaction

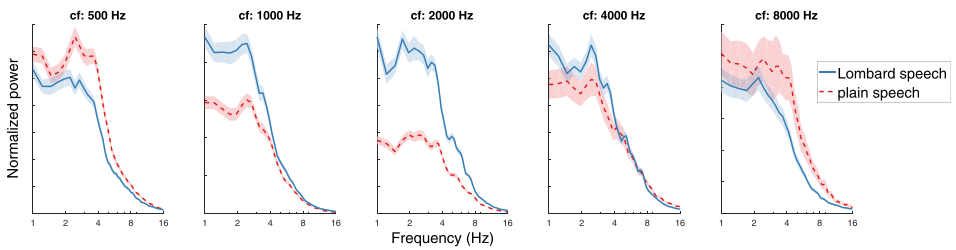

Fig. 2. (Color online) Average modulation spectra, calculated from the filterbank analysis (center frequencies at 500–8000 Hz), of Lombard speech (solid line) and matching plain speech (dashed line), averaging over the different corpora (CF = center frequency). Shaded areas indicate 95% CIs.

between Condition and the 1000 Hz frequency band showed that the effect of Condition was comparable in the 1000 and 2000 Hz frequency bands. Second, the effect of Condition was considerably smaller or even absent in the 500, 4000, and 8000 Hz frequency bands (500 Hz: $\beta = -0.004$, $SE = 0.0001$, $t = -22.815$; 4000 Hz: $\beta = -0.003$, $SE = 0.0002$, $t = -14.705$; 8000 Hz: $\beta = -0.003$, $SE = 0.0002$, $t = -16.868$).

4. Discussion

This study compared the power of amplitude modulations in the temporal envelope of Lombard speech (sentences produced in noise) and plain speech (same sentences but produced in quiet). Speech from four different corpora (various speakers, sentences types, elicitation methods; see Table 1) was analyzed by means of modulation spectra, revealing the power of frequency components in the temporal envelopes. Across all four corpora, amplitude modulations below 4 Hz were stronger in Lombard speech compared to matched plain speech (cf. similar findings for clear speech; Krause and Braida, 2002, 2004). This difference was shown to be independent of changes in speech rate and was concentrated in frequency bands around 1000 and 2000 Hz (cf. similar findings for clear speech; Krause and Braida, 2002, 2004).

The modulations most affected by Lombard speech (i.e., in delta range; 1–4 Hz) correspond to the average syllable rates (e.g., corpus 2; plain: 3.7 Hz; Lombard: 3.4 Hz). This suggests that the difference in amplitude modulations in Lombard and plain speech may be driven by more pronounced syllabic energy fluctuations in Lombard speech. The effect was concentrated in higher frequency bands (around 1000 and 2000 Hz), which is in line with studies on artificial speech enhancement. For instance, Koutsogiannaki and Stylianou (2016) found that artificially decreasing the modulation depth in lower frequency regions (200–600 Hz) and increasing the modulation depth in higher frequencies (800–3000 Hz) enhanced speech-in-noise intelligibility.

These results suggest that speakers produce more pronounced amplitude modulations in noise compared to in quiet. We interpret these findings in light of recent oscillations-based models of speech perception (Ghitza, 2011; Giraud and Poeppel, 2012; Peelle and Davis, 2012), whereby neural oscillations phase-lock (entrain) to amplitude fluctuations in speech. Greater amplitude modulations in speech produced in noise presumably help the listening brain to entrain to the attended talker, aligning neuronal excitability to the temporal structure of the attended signal, facilitating speech-in-noise perception.

This idea is corroborated by neurobiological studies showing that removing amplitude modulations, occurring at the syllabic rate, from speech reduces neural envelope-tracking activity, impairing intelligibility (Doelling *et al.*, 2014). Similarly, perception studies have shown beneficial effects of temporal modulations on phoneme identification (Drullman *et al.*, 1994) and intelligibility (Ghitza, 2012). Future neuroimaging studies should investigate whether the observed greater modulation depth in Lombard speech indeed facilitates cortical speech-tracking, aiding speech-in-noise intelligibility.

Acknowledgments

The first author was supported by a Gravitation grant from the Dutch Government to the Language in Interaction Consortium.

References and links

[1]Note that we use the terms delta, theta, and alpha to refer to specific modulation frequency bands (1–4 Hz; 4–8 Hz; 8–15 Hz, respectively) irrespective of whether these modulations occur in speech or in neural signals.

Arnal, L. H., Giraud, A.-L., and Poeppel, D. (**2015**). "A neurophysiological perspective on speech processing in 'The neurobiology of language,'" in *Neurobiology of Language*, edited by G. Hickok and S. Small (Academic Press, San Diego), pp. 463–478.

Baayen, R. H. (**2008**). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R* (Cambridge University Press, Cambridge).

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (**2013**). "Random effects structure for confirmatory hypothesis testing: Keep it maximal," J. Mem. Lang. **68**(3), 255–278.

Bosker, H. R. (**2017**). "Accounting for rate-dependent category boundary shifts in speech perception," Atten. Percept. Psychophys. **79**(1), 333–343.

Bosker, H. R., and Kösem, A. (**2017**). "An entrained rhythm's frequency, not phase, influences temporal sampling of speech," in *Proceedings of Interspeech 2017*, Stockholm.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**(3-4), 255–272.

Cooke, M., King, S., Garnier, M., and Aubanel, V. (**2014a**). "The listening talker: A review of human and algorithmic context-induced modifications of speech," Comput. Speech Lang. **28**(2), 543–571.

Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., and Tang, Y. (**2013**). "Evaluating the intelligibility benefit of speech modifications in known noise conditions," Speech Commun. **55**(4), 572–585.
Cooke, M., Mayo, C., and Villegas, J. (**2014b**). "The contribution of durational and spectral changes to the Lombard speech intelligibility benefit," J. Acoust. Soc. Am. **135**(2), 874–883.
Ding, N., Patel, A., Chen, L., Butler, H., Luo, C., and Poeppel, D. (**2017**). "Temporal modulations in speech and music," Neurosci. Biobehav. Rev. **81**, 181–187.
Ding, N., and Simon, J. Z. (**2012**). "Emergence of neural encoding of auditory objects while listening to competing speakers," Proc. Natl. Acad. Sci. U.S.A. **109**(29), 11854–11859.
Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (**2014**). "Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing," NeuroImage **85**, 761–768.
Dreher, J. J., and O'Neill, J. (**1957**). "Effects of ambient noise on speaker intelligibility for words and phrases," J. Acoust. Soc. Am. **29**(12), 1320–1323.
Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**). "ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment," Audiology **40**(3), 148–157.
Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of reducing slow temporal modulations on speech recognition," J. Acoust. Soc. Am. **95**(5), 2670–2680.
Folk, L., and Schiel, F. (**2011**). "The Lombard effect in spontaneous dialog speech," in *Proceedings of Interspeech*, Florence, pp. 2701–2704.
Garnier, M., and Henrich, N. (**2014**). "Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?," Comput. Speech Lang. **28**(2), 580–597.
Ghitza, O. (**2011**). "Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm," Front. Psychol. **2**, 130.
Ghitza, O. (**2012**). "On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum," Front. Psychol. **3**, 238.
Giraud, A.-L., and Poeppel, D. (**2012**). "Cortical oscillations and speech processing: Emerging computational principles and operations," Nat. Neurosci. **15**(4), 511–517.
Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., and Simon, J. Z. (**2013**). "Mechanisms underlying selective neuronal tracking of attended speech at a 'cocktail party,'" Neuron **77**(5), 980–991.
Golumbic, E. M. Z., Poeppel, D., and Schroeder, C. E. (**2012**). "Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective," Brain Lang. **122**(3), 151–161.
Houtgast, T., and Steeneken, H. J. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**(3), 1069–1077.
Kerlin, J. R., Shahin, A. J., and Miller, L. M. (**2010**). "Attentional gain control of ongoing cortical speech representations in a 'cocktail party,'" J. Neurosci. **30**(2), 620–628.
Kösem, A., Bosker, H. R., Takashima, A., Jensen, O., Meyer, A., and Hagoort, P. (**2017**). "Neural entrainment determines the words we hear," bioRxiv 1–22.
Koutsogiannaki, M., and Stylianou, Y. (**2016**). "Modulation enhancement of temporal envelopes for increasing speech intelligibility in noise," in *Proceedings of Interspeech*, pp. 2508–2512.
Krause, J. C., and Braida, L. D. (**2002**). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**(5), 2165–2172.
Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**(1), 362–378.
Lombard, E. (**1911**). "Le signe de l'elevation de la voix" ("The sign of the elevation of the voice"), Ann. Maladies l'Oreille Larynx **37**(2), 101–119.
Lu, Y., and Cooke, M. (**2008**). "Speech production modifications produced by competing talkers, babble, and stationary noise," J. Acoust. Soc. Am. **124**(5), 3261–3275.
Mayo, C., Aubanel, V., and Cooke, M. (**2012**). "Effect of prosodic changes on speech intelligibility," in *Proceedings of Interspeech*, Portland, pp. 1708–1711.
Peelle, J. E., and Davis, M. H. (**2012**). "Neural oscillations carry speech rhythm through to comprehension," Front. Psychol. **3**, 320.
Pittman, A. L., and Wiley, T. L. (**2001**). "Recognition of speech produced in noise," J. Speech Lang. Hear. Res. **44**(3), 487–496.
Riecke, L., Formisano, E., Sorger, B., Başkent, D., and Gaudrain, E. (**2018**). "Neural entrainment to speech modulates speech intelligibility," Curr. Biol **28**, 161–169.
Rimmele, J. M., Golumbic, E. M. Z., Schröger, E., and Poeppel, D. (**2015**). "The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene," Cortex **68**, 144–154.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**(5234), 303–304.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature **416**(6876), 87–90.
Steeneken, H. J., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**(1), 318–326.
Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**(3), 917–928.
Uchanski, R. M. (**2008**). "Clear speech," in *The Handbook of Speech Perception* (Blackwell, Hoboken, NJ), pp. 207–235.
Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., and Lorenzi, C. (**2017**). "A cross-linguistic study of speech modulation spectra," J. Acoust. Soc. Am. **142**(4), 1976–1989.
Valentino-Botinhao, C. (**2013**). http://datashare.is.ed.ac.uk/handle/10283/347 (Last viewed January 8, 2018).