

Neural Segregation of Concurrent Speech: Effects of Background Noise and Reverberation on Auditory Scene Analysis in the Ventral Cochlear Nucleus

Mark Sayles, Arkadiusz Stasiak and Ian M. Winter

Abstract Concurrent complex sounds (e.g., two voices speaking at once) are perceptually disentangled into separate “auditory objects”. This neural processing often occurs in the presence of acoustic-signal distortions from noise and reverberation (e.g., in a busy restaurant). A difference in periodicity between sounds is a strong segregation cue under quiet, anechoic conditions. However, noise and reverberation exert differential effects on speech intelligibility under “cocktail-party” listening conditions. Previous neurophysiological studies have concentrated on understanding auditory scene analysis under ideal listening conditions. Here, we examine the effects of noise and reverberation on periodicity-based neural segregation of concurrent vowels /a/ and /i/, in the responses of single units in the guinea-pig ventral cochlear nucleus (VCN): the first processing station of the auditory brain stem. In line with human psychoacoustic data, we find reverberation significantly impairs segregation when vowels have an intonated pitch contour, but not when they are spoken on a monotone. In contrast, noise impairs segregation independent of intonation pattern. These results are informative for models of speech processing under ecologically valid listening conditions, where noise and reverberation abound.

Keywords Vowels · Double vowels · Intonation · Reverberation · Background noise · Auditory scene analysis · Concurrent speech · Perceptual segregation · Inter-spike intervals · Temporal envelope · Brain stem · Cochlear nucleus

M. Sayles (✉) · A. Stasiak · I. M. Winter
Centre for the Neural Basis of Hearing, The Physiological Laboratory, Department of
Physiology, Development and Neuroscience, Downing Street, Cambridge CB2 3EG, UK
e-mail: sayles.m@gmail.com

M. Sayles
Laboratory of Auditory Neurophysiology, Campus Gasthuisberg, O&N II, Herestraat 49—bus
1021, 3000 Leuven, Belgium

© The Author(s) 2016

P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, Advances in Experimental Medicine and Biology 894,
DOI 10.1007/978-3-319-25474-6_41

1 Introduction

A difference in periodicity between simultaneous complex sounds is a strong segregation cue under quiet, anechoic conditions (Brokx and Noolboom 1982). However, noise and reverberation can both degrade speech intelligibility under realistic “cocktail-party” listening conditions (Nabelek 1993; Culling et al. 1994, 2003). Neurophysiological studies of concurrent-sound segregation have concentrated on harmonic complex sounds, with a fundamental-frequency difference (ΔF_0), heard under idealized (quiet, anechoic) conditions (e.g., Palmer 1990; Keilson et al. 1997; Larsen et al. 2008). We examine the effects of noise and reverberation, separately, on periodicity-based neural segregation of ΔF_0 concurrent vowels, with and without simulated intonation, in the ventral cochlear nucleus (VCN); the first processing station in the auditory brain stem.

2 Methods

2.1 *Animal Model*

Experiments were carried out in accordance with the United Kingdom Animals (Scientific Procedures) Act (1986), with approval of the University of Cambridge Animal Welfare Ethical Review Board. Details of our recording techniques are available elsewhere (Sayles and Winter 2008; Sayles et al. 2015). Adult guinea pigs (*Cavia porcellus*) were anesthetized with urethane and hypnorm (fentanyl/fluanisone). The cochlear nucleus was exposed via a posterior-fossa craniotomy and unilateral cerebellotomy. A glass-insulated tungsten microelectrode was advanced in the sagittal plane through the VCN, using a hydraulic microdrive. Upon isolation of a single unit, best frequency (BF) and threshold were determined. Units were classified on their responses to BF-tones.

2.2 *Complex Stimuli*

Stimuli were synthetic vowels /a/ and /i/, generated using a MATLAB implementation of the Klatt formant synthesizer. Formant frequencies were {0.7, 1.0, 2.4, 3.3} kHz for /a/, and {0.3, 2.2, 3.0, 3.7} kHz for /i/. Stimuli were presented monaurally; either alone, or as “double vowels” /a, i/ (Fig. 1). F_0 was either static (125 Hz, or 250 Hz for /a/, and 100 Hz, or 200 Hz for /i/), or sinusoidally modulated at 5 Hz, by ± 2 semitones. Reverberation was added by convolution with real-room impulse responses recorded in a long corridor, at source-receiver distances of 0.32, 2.5, and 10 m (Tony Watkins; University of Reading, UK). We refer to these as “mild”, “moderate” and “strong” reverberation. For noise-masked vowels, Gaussian noise (5-kHz bandwidth) was added at signal-to-noise ratios of {10, 3, 0} dB.

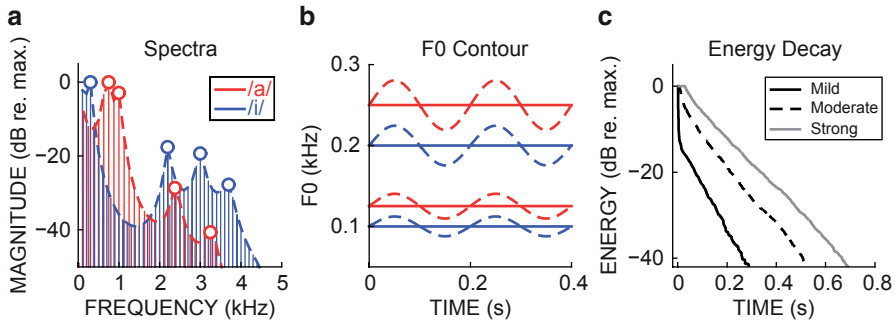


Fig. 1 Synthetic-vowel stimuli. **a** Magnitude spectra. **b** Fundamental-frequency contours: *solid lines*, static-F0 vowels; *dashed lines*, modulated-F0 vowels. **c**, Impulse-response energy-decay curves

2.3 Analyses

2.3.1 Correlograms

Analyses are based on normalized shuffled inter-spike-interval distributions derived from the VCN-unit spike trains (Fig. 2; Joris et al. 2006; Sayles et al. 2015). We computed the *across*-spike-train shuffled inter-spike-interval distribution in a short (30-ms duration) rectangular time window, slid in 5-ms steps through the 400-ms duration response. We refer to these time-varying inter-spike-interval distributions as *correlograms*.

2.3.2 Periodic Templates

We applied a “periodicity-sieve” analysis to the correlograms to estimate the dominant period(s) in the inter-spike-interval statistics (e.g., Larsen et al. 2008); yielding *template-contrast functions*. To assess the statistical significance of peaks in the template-contrast function we used a bootstrap technique (permutation analysis; 1000 replications; $p < 0.01$ considered significant). Based on the template-contrast functions in response to double vowels, we computed the “periodicity-tagged” firing rate for each vowel of the mixture (similar to that proposed by Keilson et al. 1997); e.g., for the /a/ component:

$$\bar{R}_{/a/} = \max \left(\left(\frac{\bar{R} \cdot C_{/a/}}{C_{/a/} + C_{/i/}} \right) - (0.5 \cdot \bar{R}), 0 \right)$$

Where \bar{R} is the mean firing rate of that single unit to the double vowel /a, i/, and $C_{/a/}$ and $C_{/i/}$ are template-contrast values for the two double-vowel components, respectively.

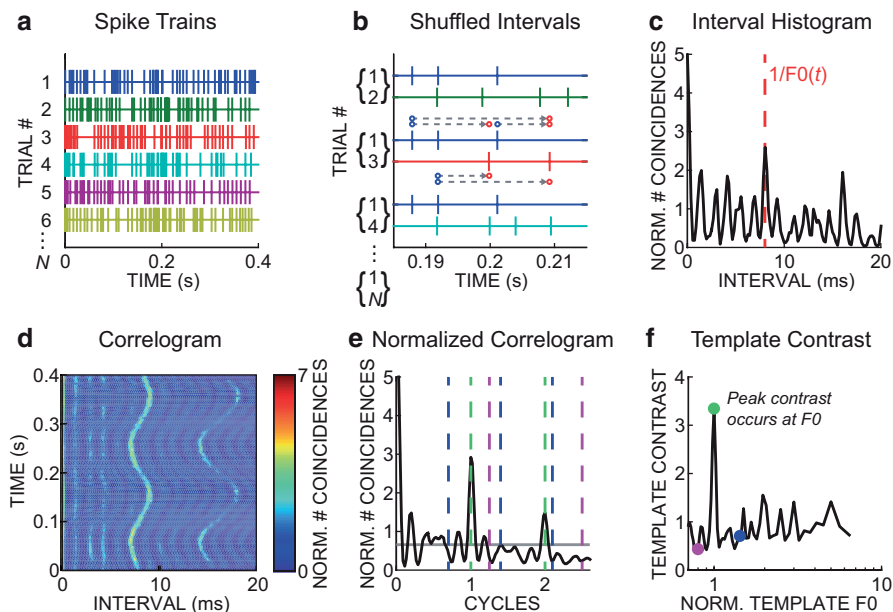


Fig. 2 Example analyses for a single unit (Primary-like, BF = 2.1 kHz), responding to a single vowel /a/ with F_0 modulated at 5 Hz (± 2 semitones) around a mean F_0 of 125 Hz. **a** Spike times were collected in response to N (typically, 50) repetitions of each stimulus. **b** Forward inter-spike intervals were calculated between all non-identical spike-train pairs, in 30-ms time windows. **c** Intervals were tallied in a histogram, and the analysis window slid in 5-ms steps to give the interval distribution as a function of time. **d**, **e** Time-varying interval distributions were normalized for instantaneous stimulus F_0 , and averaged over time. **f** Harmonic periodicity sieves were applied to compute the template-contrast function

3 Results

We recorded responses to single and double vowels, in anechoic and reverberant conditions, from 129 units with BFs between 0.1 and 6 kHz (36 primary-like [PL/PN], 47 chopper [CT/CS], 24 onset [OC/OL], 19 low-frequency [LF], 3 unusual [UN]). From 52 of these, we also recorded responses to vowels in noise. The effects of noise and reverberation on double-vowel segregation are not equivalent. The results can be summarized as: (1), There is a strong interaction between F_0 modulation (simulating intonated speech) and reverberation to reduce template contrast. (2), There is no interaction between F_0 modulation and signal-to-noise ratio. (3), Noise-induced deficits in neural periodicity-based double-vowel segregation are strongly BF-dependent, due to more total (masking) noise power passed by higher-BF filters. (4), Reverberation impairs neural segregation of intonated double vowels independent of BF, but has only marginally detrimental effects on segregation of double vowels with steady F_0 s.

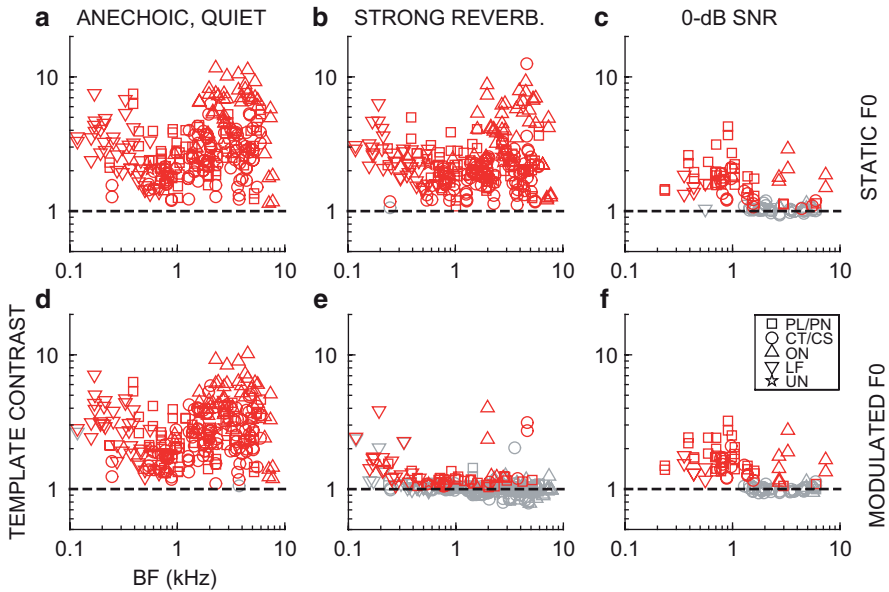


Fig. 3 Template contrast vs. unit BF for responses to the single vowel /a/. *Top row*, responses to static-F0 vowels; *Bottom row*, responses to modulated-F0 vowels. Different symbols indicate unit type (legend in F). *Red symbols* indicate significant values ($p < 0.01$), *gray symbols* are non-significant. *Dashed line* indicates contrast of 1, the expected value for a flat correlogram

Figure 3 shows the template contrast calculated from responses to the single vowel /a/ under ideal (anechoic, quiet) conditions, in the presence of “strong” reverberation, and in the presence of background noise at 0-dB SNR. There is a small reduction in template contrast due to reverberation alone (Fig. 3a, b), and a similar small reduction due to F0 modulation alone (Fig. 3a, d). However, the major detrimental effect on template contrast is an interaction between reverberation and F0 modulation (Fig. 3a, e). In the combined presence of reverberation and F0 modulation many units do not have a significant representation of the periodicity corresponding to this single vowel in their inter-spike-interval distributions; the many non-significant gray symbols in Fig. 3e. In contrast, the presence of masking noise reduces template contrast at the F0 period independent of F0 modulation, but in a BF-dependent manner with higher-BF units affected most strongly (Fig. 3c, f).

The interaction between reverberation and F0 modulation is further illustrated by plotting template contrast in response to modulated-F0 vowels against that in response to static-F0 vowels (Fig. 4). For both single- and double-vowel responses, the reduction in template contrast in the presence of reverberation is much greater in the modulated-F0 condition relative to the static-F0 condition (Fig. 4b, e). In the presence of noise, the maximum template contrast is reduced equally for modulated- and static-F0 vowels (Fig. 4c, f).

Based on template contrast calculated from double-vowel responses, we computed a “periodicity-tagged” discharge rate for each vowel of the concurrent-speech

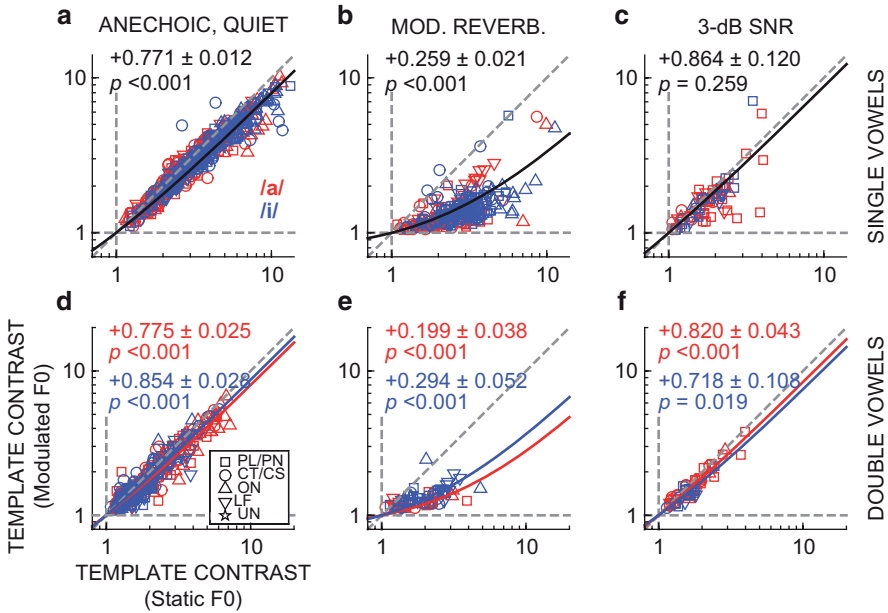


Fig. 4 Template contrast: modulated-F0 vs. static-F0 vowels. *Top row*, responses to single vowels; *Bottom row*, responses to double vowels. Red symbols indicate responses to /a/, and blue symbols /i/. Horizontal and vertical gray dashed lines indicate template contrast of 1. Diagonal gray dashed line indicates equality. Solid lines indicate least-squares linear fits to the significant data, with the fit constrained to (1,1). Text in each panel indicates the linear-regression slope (β_1) \pm S.E., and the *p*-value for a two-tailed *t*-test with the null hypothesis $H_0: \beta_1=1$. For clarity, only those data from responses with significant contrast ($p<0.01$) for both static *and* modulated vowels are displayed

mixture (Fig. 5). In the across-BF profiles of periodicity-tagged rate, there are peaks and troughs corresponding to the formant structure of each vowel under quiet, anechoic listening conditions. This is the case with both static- and modulated-F0 double vowels (Fig. 5a, d). With reverberation, there are no formant related peaks and troughs remaining in the modulated-F0 case (Fig. 5e), although a clear formant-related pattern of periodicity-tagged discharge rate remains in the static-F0 case in reverberation (Fig. 5b). Information about higher formants is degraded in the presence of noise; however, formant-related peaks in periodicity-tagged discharge rate remain in the first-formant region for both static- and modulated-F0 vowels (Fig. 5c, f).

To quantify the “segregation” of the two vowels of the double-vowel mixture by the periodicity-tagged firing-rate profile across BF, we computed the normalized Euclidean distance between the /a/ and /i/ profiles in response to /a, i/ as (Fig. 6),

$$d_{(a/,i/)} = \frac{\sum_{j=1}^n (\bar{R}_{/a/,j} - \bar{R}_{/i/,j})^2}{\sqrt{n}}$$

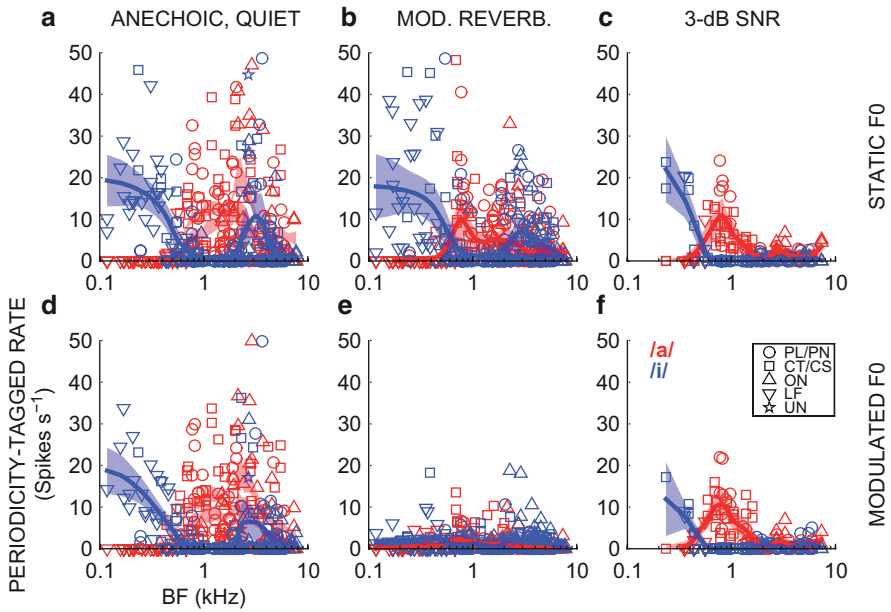


Fig. 5 Periodicity-tagged firing rate vs. BF. *Top row*, responses to static-F0 vowels; *Bottom row*, responses to modulated-F0 vowels. *Solid lines*, lowess smoothing fits; *shaded areas*, 95% confidence intervals

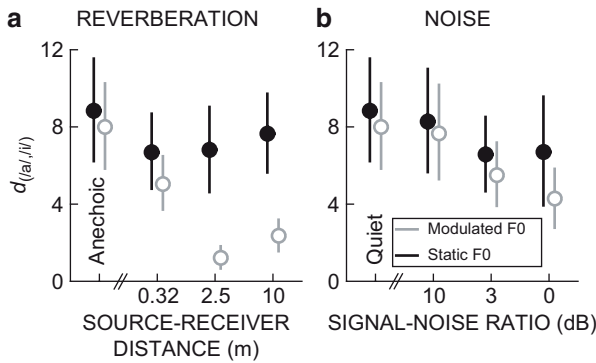


Fig. 6 Euclidean distance between /a/ and /i/ periodicity-tagged spike-rate vs. BF profiles, calculated from the double-vowel responses

The double-vowel sounds are well segregated on the basis of the periodicity-tagged discharge rate statistic, except when F0 is modulated and the sounds are heard in “moderate” or “strong” reverberation (Fig. 6a).

4 Discussion

The physical effects of noise and reverberation have some similarities: small changes to the magnitude and phase of each component of a complex sound. The impulse response of a room is a linear filter. Therefore, for a steady-state source no new frequency components are added to a reverberant acoustic signal. Contrast this with additive noise. Each indirect sound component in a reverberant space adds to the direct sound at the receiver with essentially random phase, reducing the depth of temporal-envelope modulation at the output of cochlear band-pass filters (Sabine 1922; Sayles and Winter 2008; Sayles et al. 2015; Slama and Delgutte 2015). Perceptually, noise and reverberation can both decrease speech intelligibility (e.g., Nabelek 1993; Payton et al. 1994), and can be particularly troublesome for cochlear-implant listeners (e.g., Qin and Oxenham 2005). However, error patterns differ substantially between noisy and reverberant spaces (Nabelek 1993; Assmann and Summerfield 2004). Moreover, normal-hearing listeners have remarkably good speech understanding in moderately reverberant spaces (e.g., Poissant et al. 2006).

Our results can be understood in terms of the effects of reverberation on the acoustic temporal envelope. For static-F0 vowels, the reduction in envelope modulation due to reverberant energy is evident in higher-BF units (Fig. 3a, b; Fig. 5a, b). The addition of frequency modulation to the vowels has a dramatic effect on template contrast in reverberation, due to spectral smearing: i.e., monaural decorrelation. Noise impairs neural segregation of concurrent vowels independent of intonation pattern, but in a BF-dependent manner. Similar to other studies of noise-masked single-vowel coding (e.g., Delgutte and Kiang 1984; May et al. 1998), this is the consequence of higher-BF filters passing more total noise power than low-BF filters. The within-band signal-to-noise ratio is therefore much lower for high-BF units. The negative spectral tilt of the vowels, masked by flat-spectrum noise, imposes a similar frequency-dependency on signal-to-noise ratio after cochlear filtering.

The differential effects of noise and reverberation on speech representations in the early stages of brain-stem neural processing are in clear contrast to noise- and reverberation-invariant representations of speech in auditory cortex (Mesgarani et al. 2014). Perhaps one clue to the neural underpinnings of robust speech understanding in challenging acoustic environments is compensation for the effects of reverberation on coding of temporal-envelope modulation in the inferior colliculus, based on sensitivity to inter-aural correlation (Slama and Delgutte 2015). What is clear from our data is that neurophysiological effects of room reverberation cannot simply be assimilated to those of broadband noise for signal detection and discrimination.

Acknowledgements Funded by a grant from the BBSRC (IMW), and a University of Cambridge MD/PhD scholarship (MS).

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Assmann PF, Summerfield AQ (2004) The perception of speech under adverse conditions. In: Greenberg S, Ainsworth WA, Fay RR, Popper A (eds) *Speech processing in the auditory system*. Springer, New York
- Brokx JPL, Nooteboom SG (1982) Intonation and the perceptual separation of simultaneous voices. *J Phon* 10(1):23–36
- Culling JF, Summerfield Q, Marshall DH (1994) Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels. *Speech Comm* 14(1):71–95
- Culling JF, Hodder KI, Toh CY (2003) Effects of reverberation on perceptual segregation of competing voices. *J Acoust Soc Am* 114(5):2871–2876
- Delgutte B, Kiang NYS (1984) Speech coding in the auditory-nerve. V. vowels in background-noise. *J Acoust Soc Am* 75(3):908–918
- Joris PX, Louage DH, Cardeon L, van der Heijden M (2006). Correlation index: a new metric to quantify temporal coding. *Hear Res*, 216–217(1), 19–30
- Keilson SE, Richards VM, Wyman BE, Young ED (1997) The representation of concurrent vowels in the cat anaesthetized ventral cochlear nucleus: evidence for a periodicity-tagged spectral representation. *J Acoust Soc Am* 102(2):1056–1070
- Larsen E, Cedolin L, Delgutte B (2008) Pitch representations in the auditory nerve: two concurrent complex tones. *J Neurophysiol* 100(3):1301–1319
- May BJ, Prell GS, Sachs MB (1998) Vowel representations in the ventral cochlear nucleus of the cat: effects of level, background noise, and behavioral state. *J Neurophysiol* 79(4):1755–1767
- Mesgarani N, David SV, Fritz JB, Shamma SA (2014) Mechanisms of noise robust representation of speech in primary auditory cortex. *Proc Natl Acad Sci USA* 111(18):6792–6797
- Nabelek AK (1993) Communication in noisy and reverberant environments. In: Stuebelaker GA, Hochberg I (eds) *Acoustical factors affecting hearing aid performance*. Allyn and Bacon, Needham Heights
- Palmer AR (1990) The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers. *J Acoust Soc Am* 88(3):1412–1426
- Payton KL, Uchanski RM, Braida LD (1994) Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *J Acoust Soc Am* 95(3):1581–1592
- Poissant SF, Whitmal NA 3rd, Freyman RL (2006) Effects of reverberation and masking on speech intelligibility in cochlear implant simulations. *J Acoust Soc Am* 119(3):1606–1615
- Qin MK, Oxenham AJ (2005) Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear Hear* 26:451–460
- Sabine WC (1922) *Collected papers on acoustics*. Harvard University Press, Cambridge
- Sayles M, Winter IM (2008) Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron* 58:789–801
- Sayles M, Stasiak A, Winter IM (2015) Reverberation impairs brainstem temporal representations of voiced vowel sounds: challenging “periodicity-tagged” segregation of competing speech in rooms. *Front Syst Neurosci* 8:248
- Slama MC, Delgutte B (2015) Neural coding of sound envelope in reverberant environments. *J Neurosci* 35(10):4452–4468