Hearing Research 341 (2016) 168–178

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Speech perception adjusts to stable spectrotemporal properties of the listening environment

Christian E. Stilp [a, *], Paul W. Anderson [b], Ashley A. Assgari [a], Gregory M. Ellis [a], Pavel Zahorik [a]

[a] Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY, 40292, USA
[b] Department of Psychology, Murray State University, Murray, KY, 42071, USA

ARTICLE INFO

Article history:
Received 15 March 2016
Received in revised form
10 August 2016
Accepted 15 August 2016
Available online 3 September 2016

Keywords:
Speech perception
Spectral calibration
Perceptual constancy
Reverberation

ABSTRACT

When perceiving speech, listeners compensate for reverberation and stable spectral peaks in the speech signal. Despite natural listening conditions usually adding both reverberation and spectral coloration, these processes have only been studied separately. Reverberation smears spectral peaks across time, which is predicted to increase listeners' compensation for these peaks. This prediction was tested using sentences presented with or without a simulated reverberant sound field. All sentences had a stable spectral peak (added by amplifying frequencies matching the second formant frequency [F_2] in the target vowel) before a test vowel varying from /i/ to /u/ in F_2 and spectral envelope (tilt). In Experiment 1, listeners demonstrated increased compensation (larger decrease in F_2 weights and larger increase in spectral tilt weights for identifying the target vowel) in reverberant speech than in nonreverberant speech. In Experiment 2, increased compensation was shown not to be due to reverberation tails. In Experiment 3, adding a pure tone to nonreverberant speech at the target vowel's F_2 frequency increased compensation, revealing that these effects are not specific to reverberation. Results suggest that perceptual adjustment to stable spectral peaks in the listening environment is not affected by their source or cause.

1. Introduction

Much in the sensory environment is predictable from time to time and place to place. Sensory systems have adapted and evolved to be sensitive to this predictability (Attneave, 1954; Barlow, 1961). Sensitivity to stable aspects of the environment promotes a host of perceptual phenomena: grouping, scene analysis, and source localization, among others. The auditory system responds to predictability in the sensory environment through several related mechanisms: adaptation, constancy, normalization, compensation, and calibration. While some of these mechanisms may differ only in name or in scale, they all serve audition by providing adjustment to stable aspects of the listening environment. Compensating for regularities in the sensory environment can affect processing of simple acoustic properties, such as adapting to a particular frequency or entraining to a regular rhythm. Environmental regularities can also affect higher-level auditory processing, such as sound source identification, speech understanding, and object recognition. The present focus is on stable spectrotemporal properties of the listening environment that influence speech perception.

In everyday perception, sounds are filtered by the listening environment. As sounds propagate from source to perceiver, different frequencies are amplified or attenuated depending on the composition of the listening environment. This filtering can make certain frequencies particularly prominent and relatively stable across time, producing *stable spectral properties*. Several reports have shown that perception deemphasizes these stable properties and increases reliance on changing (less predictable, and thus more informative) signal properties (Kiefte and Kluender, 2008; Alexander and Kluender, 2010; Stilp and Anderson, 2014). For example, when earlier sounds feature a stable spectral peak that matches the second formant frequency (F_2) of the following target vowel (perceptually varying from /i/ to /u/, for which F_2 is a key distinguishing feature), listeners decreased their reliance on F_2 and

* Corresponding author.
E-mail addresses: christian.stilp@louisville.edu (C.E. Stilp), panderson9@murraystate.edu (P.W. Anderson), ashley.assgari@louisville.edu (A.A. Assgari), g.ellis@louisville.edu (G.M. Ellis), pavel.zahorik@louisville.edu (P. Zahorik).

http://dx.doi.org/10.1016/j.heares.2016.08.004
0378-5955/

increased their reliance on changing, more informative cues for vowel identification such as spectral tilt. In the literature, this process has been called *auditory perceptual calibration* (Kiefte and Kluender, 2008; Alexander and Kluender, 2010; Stilp and Anderson, 2014). Here we adopt the more descriptive term *spectral calibration* to distinguish it from calibration to other stable acoustic properties or contingencies between properties. Spectral calibration is a key mechanism for factoring out predictable acoustic aspects of the listening environment, and is analogous to color constancy in vision (see Alexander and Kluender, 2010; Stilp et al., 2010 for discussions).

Natural listening environments also produce acoustic reflections that interact with the source signal. Reverberant acoustic energy can degrade the intelligibility and quality of speech, especially when reverberation times are long (Knudsen, 1929; Nábělek and Robinson, 1982; Nábělek and Donahue, 1984; Nábělek and Letowski, 1985). However, given sufficient exposure, listeners compensate for stable patterns of reverberation (Watkins, 2005a, 2005b; Watkins et al., 2011; Watkins and Makin, 2007; Brandewie and Zahorik, 2010, 2013; Srinivisan and Zahorik, 2013, 2014). This compensation has been shown to improve speech intelligibility considerably (Brandewie and Zahorik, 2010, 2013; Srinivisan and Zahorik, 2013, 2014). From this perspective, compensating for reverberation is another instance of perceptual constancy in speech perception (Assmann and Summerfield, 2004; Watkins and Makin, 2007). While not traditionally viewed as such, a given pattern of reverberation can serve as a *stable spectrotemporal property* of the listening environment, producing characteristic spectrotemporal alterations to the source signal.

Listening environments alter the frequency compositions of sounds while also producing acoustic reflections. Listeners often factor out these stable properties of the acoustic environment, whether they are primarily spectral (as in spectral calibration) or spectrotemporal (as in compensation for reverberation). In terms of perceptual adjustment to stable properties of the listening environment, compensations for stable spectral properties and reverberation are highly related. Intriguingly, these processes offer separate notions of what makes a particular spectral property "stable", whether it is prominence in the spectral domain (*e.g.*, relative amplitude of a spectral peak, stability of overall spectral shape) or the spectrotemporal domain (*e.g.*, patterns of temporal elongation due to acoustic reflections).

Despite their broad similarities and likely co-occurrence in speech perception, spectral calibration and compensation for reverberation have been studied separately using different tasks. Compensation for reverberation has been studied by measuring speech intelligibility or word recognition, while spectral calibration studies have examined the perceptual weighting of spectral cues for vowel categorization. While speech intelligibility and cue weighting are not unrelated (Winn and Litovsky, 2015), they are sufficiently distinct to obscure the relative contributions of spectral calibration and compensation for reverberation to speech perception.

In studies of spectral calibration, stable spectral peaks were added to a preceding acoustic context through filtering. This filtered context featured a stable spectral peak that matched the F_2 center frequency in the subsequent target vowel, and identification of this vowel was altered by this spectral peak in earlier sounds. Previous investigations used filters with narrow bandwidths (100 Hz; Fig. 1a) in order to minimally affect speech quality and intelligibility. However, reverberation can significantly impair speech quality in ways that produce clear predictions for how spectral cue use would differ in reverberant and nonreverberant listening conditions. Reverberation acts as a low-pass filter in the amplitude modulation domain (Houtgast and Steeneken, 1973). Stable spectral peaks wax and wane along with speech energy in the passband region (generally between 3 and 8 Hz; Houtgast and Steeneken, 1985; Elliott and Theunissen, 2009) and would be smeared across time in reverberant listening conditions (Fig. 1b). Hearing stable spectral properties more often in a fixed amount of

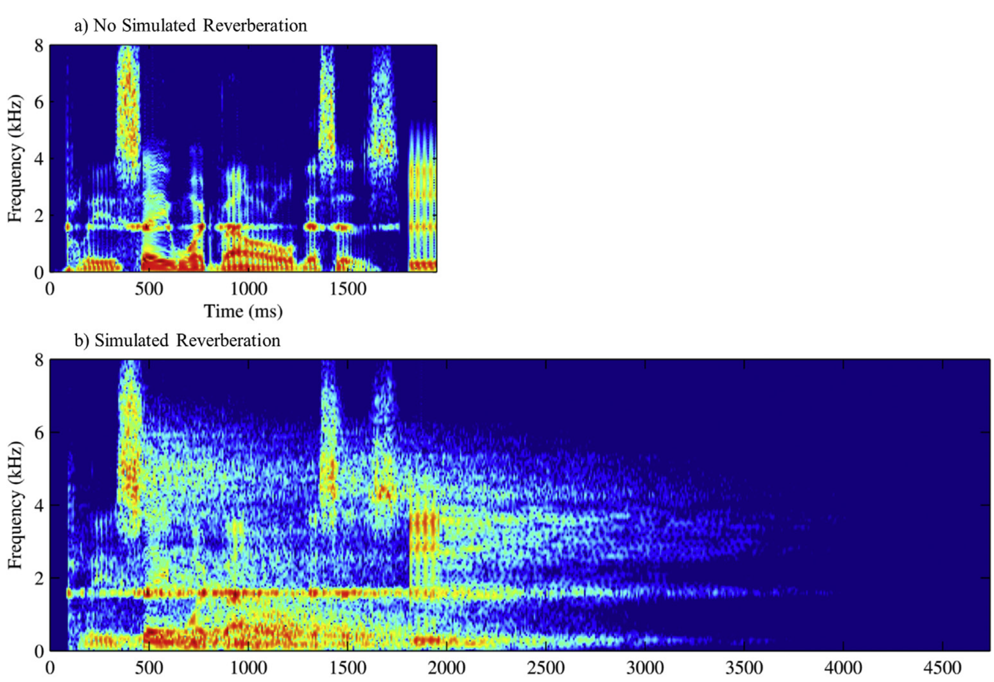

Fig. 1. Spectrograms for sample trials from Experiment 1. Both precursor sentences ("Please say what vowel this is") featured a stable spectral peak at 1600 Hz, followed by a target vowel with $F_2 = 1600$ Hz and spectral tilt $= -3$ dB/octave. (a) Sentence without simulated reverberation, Experiment 1a. (b) Sentence in a simulated reverberant sound field (broadband $T_{60} = 3476$ ms), Experiment 1b. Spectrograms are time-aligned to illustrate longer trial durations for reverberant stimuli.