Journal of Phonetics 51 (2015) 1–5

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

High is not just the opposite of Low☆

Jonathan Paul Evans

Academia Sinica, Taiwan

ARTICLE INFO

Article history:
Received 29 April 2015
Received in revised form
9 May 2015
Accepted 14 May 2015
Available online 17 June 2015

Keywords:
Fundamental frequency
High
Pitch
Tone
Stress
Larynx
Music

ABSTRACT

Elevated fundamental frequency (*F*0) has been found to have similar properties across languages. For example, raised pitch often accompanies syllabic stress, emotionally charged speech, infant-directed speech, and questions. In many languages, occurrence of high tone is subject to more constraints than are other tones. Given that these patterns occur frequently in the world's languages, it is natural to ask whether language-independent properties of raised *F*0 could play a role in the existence of typological similarities. The purpose of the present paper is to survey possible language-external factors that appear to play a role in the special status of linguistic H(igh). Moreover, the collection of studies assembled in this Special Issue provides empirical evidence that raised *F*0 attracts listener attention differently from lowered *F*0, that sustained production of high *F*0 may involve unique auditory control mechanisms, and that social context and even semantics may contribute to speaker production of raised *F*0. It is hoped that the articles of the special issue will provide a phonetic basis to explain some of the asymmetries observable in prosodic systems of languages around the world.

1. Introduction

On a hot summer's day, director Mel Larimer began to prepare the Interlochen National Music Camp high school choir to learn the choral portion of Beethoven's Ninth Symphony. He asked the pianist to play high A (880 Hz), the recurring highest note sung more than seventy times throughout the piece. He then asked those first sopranos who were able, to sing the note, which audibly strained their voices. Lastly, he asked the choir, “Which sounded more impressive, the piano or the human voice?” That moment in 1984 planted the seed that has become this article and this Special Issue.

This article sets out to introduce findings that show that raising human voice fundamental frequency is not the mirror image of lowering it. That is, evidence from production and perception suggest that there are physical and psychological bases for the widespread linguistic asymmetries between H(igh) and L(ow).

Because production and perception issues are discussed below and in the articles in this Special Issue, I briefly mention here some common phonological processes that demonstrate a privileged position for elevated pitch. In many languages with H and L tones, there is an active constraint against adjacent H tones (“Meussen's rule” Goldsmith, 1984). Even more restrictive, some phonological grammars do not permit more than one H tone per word (“culminativity”; Evans, 2008). Tonal systems with similar restrictions against L are much less common (Hyman, 2001). In addition, in many cases, raised *F*0 on an individual syllable corresponds with stress (Crystal, 2011).

It is not clear whether acoustic properties of sound play a direct role in the special characteristics of linguistic H. Sound intensity is a property of amplitude, rather than frequency. Thus, a sound with higher *F*0 does not contain more energy than one with lower *F*0. Nevertheless, studies of *F*0 and loudness, the perceptual correlate of intensity, have shown that different fundamental frequencies are perceived at different loudness, even with sound pressure level kept constant (Fletcher & Munson, 1933; Robinson & Dadson, 1956). However, these studies tested responses to pure sinusoidal tones, not complex voice-like tones; they also did not probe distinctions within the normal human spoken *F*0 range. Thus, it is still unknown what shape a vocal pitch/loudness curve would take.

☆For their comments on earlier versions of this article, I would like to thank Ken DeJong, Yu-Wen Lai and Wen-Chi Yeh. Others are acknowledged within the article for their domain-specific assistance. Funding for this research was provided by Ministry of Science and Technology (Taiwan, R.O.C.) Grant #100-2628-H-001-008-MY4.

E-mail address: jonathan@sinica.edu.tw

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.05.001

Fundamental frequency is conveyed multi-modally in the sound signal. It is the lowest common denominator of the component frequencies, because the harmonic frequencies are multiples of the vibration rate of the vocal folds. Thus, *F*0 is also the distance between each pair of neighboring harmonics. As mentioned in the discussion on perception (Section 3), this multi-modality (wave frequency and distance between harmonics) allows multiple perception mechanisms to identify fundamental frequency.

As the sound wave travels, intensity degrades rapidly, while frequency does not. The idealized speech signal emanates from the speaker as a sphere whose surface area equals $4\pi r^2$. Thus, the sound pressure level on the surface decreases relative to the square of the distance from the source. On the other hand, wave frequency information does not degrade across distance as quickly as intensity does. As a sound wave travels through air, the wave moves air molecules. With each push of a molecule, some wave energy is lost in the form of heat. Lower frequency waves, such as *F*0, lose less energy to heat than do higher frequency waves, such as spectral components (Yun-Feng Hsieh, p.c.). Thus, the intensity of the greatest common denominator frequency decreases more slowly than does that of formants. Due to the physical properties of the sound signal, and the multi-modality of fundamental frequency, *F*0 degrades more slowly across distance than formants and intensity do.

Because *F*0 is more robust across distance than other components of the sound wave are, it is available to speakers and listeners as a relatively invariant signal carrier. Given the relationship between increased intensity and increased *F*0 (Section 2), raised *F*0 could serve as a more faithful indicator of raised intensity than the actual sound pressure level. One example of a correlation between speech intensity and raised *F*0 is that of expression of intense emotions. Feelings such as happiness, anger, fear, and even impatience tend to be expressed by sustained raised *F*0 (Michaud, Vaissière, & Nguyễn, 2015; Pell, Monetta, Paulmann, & Kotz, 2009; Schröder, 2001; Williams & Stevens, 1972). To the extent that this raised *F*0 correlates with higher subglottal pressure, it is a robust indicator of forceful speech.

The following two sections introduce findings related to production and perception of *F*0; they also highlight phenomena that suggest a special role for *F*0 raising.

2. Production and raising of *F*0

Acoustically defined, fundamental frequency (*F*0) is the “lowest frequency component in a complex sound wave” (Crystal, 2011). From a physiological perspective, *F*0 is the rate of vibration of the vocal folds (Gick, Wilson, & Derrick, 2013:86; Ryalls & Behrens, 2000:20). When speakers laryngeally raise the vibration rate, the cricothyroid muscle contracts, which tilts the thyroid cartilage forward, elongating and thinning the vocal folds. Simultaneously, the thyroarytenoid muscles contract, stiffening the vocal folds. Both actions serve to raise the frequency of vibration. In lowering *F*0, parts of the thyroarytenoid muscles contract, shortening the vocal folds. The concomitant increase of mass per unit length slows vocal fold vibration (Gick et al., 2013:86–89; Hirose, 1997:116–136; Reetz & Jongman, 2009:69–71).

Speakers monitor and adjust *F*0. Control of laryngeal structures during speech involves different neural pathways than are invoked during less volitional activities such as cough, swallow, and sniff. (Ludlow, 2005). During swallowing, coughing, etc., the muscles activated are consistent across instances and speakers. However, speakers vary between and within themselves in the mixture of subglottal pressure, airflow, and cricothyroid and thyroarytenoid muscle activation employed to yield a particular combination of intensity and *F*0 (Atkinson, 1976). Speakers rely on both somatosensory feedback and auditory feedback in order to monitor and adjust *F*0 production. Laryngeal muscles move with speed and precision during utterances, which suggests that throughout the language acquisition process, somatosensory feedback aids the speaker in producing the laryngeal gestures that yield the desired vocal output. Recent experiments suggest that mechanoreceptors in the laryngeal mucosa provide the central nervous system with feedback when the larynx is in motion (Ludlow, 2005).

Speakers monitor their own *F*0 auditorily, so that when presented with an *F*0-shifted version of their ongoing speech, they produce a compensatory shift of *F*0 in the opposite direction (Larson, White, Freedland, & Burnett, 1996; Sturgeon, Hubbard, Schmidt, & Loucks, 2015; Ning, Loucks, & Shih, 2015). Differences in compensatory shift have been noted among trained vocalists, speakers of tonal languages, non-tonal language speakers, and L2 speakers of a tonal language (Ning et al., 2015). The existence of these differences suggests that compensatory *F*0 shift is not merely a reflex. Musicians who are not vocalists differ from non-musically trained speakers in their pitch shift responses, suggesting that pitch control experience of a non-laryngeal nature affects the vocal *F*0 control mechanism (Sturgeon et al., 2015).

In addition to raising *F*0 via laryngeal settings, speech uttered with more force also has higher *F*0. For example, *F*0 increases when the rate of airflow across the vocal folds increases and all laryngeal settings are held constant, as confirmed by studies both on humans and on excised canine larynges (Alipour & Scherer, 2007; Baer, 1979; Lieberman, Knudson, & Mead, 1969; Titze, 1989). Higher subglottal pressure leading to greater airflow can occur in various contexts, both linguistic and environmental. *F*0 raising has been noted in at least four contexts in which higher subglottal pressure or more forceful speech could be a cause.

First, higher subglottal pressure could be related to the sudden rises in *F*0 noted at the beginning of new discourse sections (Menn & Boyce, 1982; Pierrehumbert & Hirschberg, 1990 (chap. 14); Mohler & Mayer, 2001; Tseng, Pin, Lee, Wang, & Chen, 2005; Tseng, 2008; Xu, 2006). Second, in the presence of noise, speakers produce speech with both greater intensity and higher *F*0; that is, the “Lombard effect” (Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Third, in addition to noise, distance from listener is also a condition in which more speech effort is used. Shih and Lu (2015) find that as the distance between talker and listener increases, there are concomitant increases in intensity, duration, *F*0 maximum and *F*0 range. An increase in *F*0 range indicates that higher