FORUM

Forum is intended for communications that raise acoustical concerns, express acoustical viewpoints, or stimulate acoustical research and applications without necessarily including new findings. Publication will occur on a selective basis when such communications have particular relevance, importance, or interest to the acoustical community or the Society. Submit such items to an appropriate associate editor or to the Editor-in-Chief, labeled FORUM. Condensation or other editorial changes may be requested of the author.

Opinions expressed are those of the individual authors and are not necessarily endorsed by the Acoustical Society of America.

Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization

Ingo R. Titze[a)]
National Center for Voice and Speech, 136 South Main Street, Suite 320, Salt Lake City, Utah 84101, USA

Ronald J. Baken
Department of Otolaryngology, New York Eye and Ear Infirmary of Mount Sinai, 310 East 14th Street, New York, New York 10003, USA

Kenneth W. Bozeman
Lawrence University Conservatory of Music, 711 East Boldt Way, Appleton, Wisconsin 54911, USA

Svante Granqvist
Royal Institute of Technology (KTH), School of Technology and Health (STH), Basic Science and Biomedicine, Campus Haninge, SE-136 40 Handen, Sweden

Nathalie Henrich
GIPSA-lab, Département Parole et Cognition, Domaine Universitaire, 11 rue des Mathématiques, BP 46 38402 Saint Martin d'Hères cedex, France

Christian T. Herbst
Laboratory of Biophysics, Department of Experimental Physics, Palacký University Olomouc, Czech Republic

David M. Howard
Department of Electronics, University of York, Heslington, York, United Kingdom

Eric J. Hunter
Communicative Sciences and Disorders, Michigan State University, 1026 Red Cedar Road, East Lansing, Michigan 48824, USA

Dean Kaelin
Dean Kaelin Vocal Studio, 2539 East 4430 Street, Holladay, Utah 84124, USA

Raymond D. Kent
The Waisman Center, University of Wisconsin–Madison, 1500 Highland Avenue, Madison, Wisconsin 53705, USA

Jody Kreiman
Head and Neck Surgery, University of California–Los Angeles School of Medicine, 1000 Veteran Avenue, Los Angeles, California 90095, USA

Malte Kob
Musikalische Akustik and Theorie der Musikübertragung Erich-Thienhaus-Institut Hochschule für Musik Detmold, Neustadt 22, 32756 Detmold, Germany

Anders Löfqvist
Haskins Laboratories, Yale University, 300 George Street, Suite 900, New Haven, Connecticut 06511, USA

Scott McCoy
School of Music, Ohio State University, 304 Mershon Auditorium, 1871 North High Street, Columbus, Ohio 43201, USA

Donald G. Miller
Voce Vista, Achterste Kamp 9, 9301 RB Roden, the Netherlands

Hubert Noé
Klostergasse 30, 8280 Fürstenfeld, Austria

Ronald C. Scherer
Communication Sciences and Disorders, Bowling Green State University, 200 Health and Human Services Building, Bowling Green, Ohio 43403, USA

John R. Smith
School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia

Brad H. Story
Speech, Language, and Hearing Sciences, University of Arizona, P.O. Box 210071, Tucson, Arizona 85721, USA

Jan G. Švec
Department of Biophysics, Palacký University, Olomouc, the Czech Republic

Sten Ternström
Department of Speech, Music and Hearing (KTH), Lindstedtsvägen 24, S-100 44 Stockholm, Sweden

Joe Wolfe
School of Physics, University of New South Wales, Sydney 2052, Australia

(Received 6 February 2015; accepted 17 April 2015)
[http://dx.doi.org/10.1121/1.4919349]

I. INTRODUCTION

The study of vocalization brings together a long history of voice terminology from acoustics, linguistics, phonetics, speech pathology, laryngology, music, theater, biology, and speech technology. One challenge is to maintain consistency in symbolic representation of key variables used for resonant frequencies of the airways and the frequencies produced by sound sources. Scientists who use mathematical notation are encouraged to use single letters with subscripts for algebraic clarity (Cohen and Giacomo, 1987), whereas clinicians often prefer multiple-letter abbreviations without subscripts for ease of written and spoken communication. For example, the symbol f_o as the fundamental frequency of oscillation of the vocal folds has been used in thousands of publications, both with upper case and lower case letters, and both with subscript and no subscript. If capitalized, the symbol is not clearly dissociated from formant frequencies $F_1, F_2, \ldots, F_n$. The subscript of fundamental frequency, if written as a "zero," does not indicate a first harmonic, but rather a meaningless "zero" harmonic. If written as an alphabetic "o," it can stand for "oscillation," which is more meaningful. Some investigators have expressed a desire to abandon f_o altogether, but such a dramatic shift would render a disservice to volumes of historic literature.

A new attempt at clarity has arisen, in which authors are beginning to identify harmonics of the sound source as $H_1, H_2, H_3, \ldots, H_n$. The problem with this notation is that the symbol "H" can refer to either the frequency or

[a)]Also at: Department of Communication Sciences and Disorders, the University of Iowa, Iowa City, IA. Electronic mail: ingo.titze@ncvs.org

the amplitude of a harmonic. For example, when an H_2/H_1 ratio is computed for a spectral balance measure, an amplitude ratio is intended, not a frequency ratio. In other cases, when authors refer to H_1/F_1 or H_1–F_1 relations (especially in singing), they are talking about frequency ratios or differences. Thus, confusion has not been eliminated by introducing the H symbol. We will show that its use is not necessary.

With regard to airway resonances, historical precedence and current usage of terminology are also slightly at odds. Joe Wolfe and colleagues suggest that the symbol R be used to stand separate from the symbol F for formant (Wolfe, 2014). The distinction is being made because a formant was originally defined as a peak in the output spectrum envelope radiated from the mouth (Hermann, 1894, 1895; Russell, 1929; Fant, 1960, p. 20). A similar definition appears in the current ASA standard of acoustic terminology (Acoustical Society of America, 2004), namely, that a formant is "a range of frequencies in which there is absolute or relative maximum in the sound spectrum. The frequency at the maximum is the formant frequency." As such, a formant involves both the source and the filter. However, as speech analysis and synthesis have progressed in a half century, the definition has not been universally maintained. Fant (1960, pp. 20, 53) defined formants as the poles of the transfer function of the supraglottal vocal tract, and labeled the pole frequencies $F_1, \ldots, F_n$ and their bandwidths $B_1, \ldots, B_n$. He was followed in this path by many authors, such as Titze (1994, p. 156) or Stevens (1998, p.131). It is noteworthy that Flanagan (1965, p. 57) was aware of the dual definition (and possible evolution) by using the term "formant resonance." While Benade (1976) maintained the definition of "peaks in the spectral envelope of the radiated sound," Badin and Fant (1984) computed formant frequencies and bandwidths on the basis of x-ray area function *resonances* of the supraglottal vocal tract, not peaks in the output spectrum envelope. Story *et al.* (1996) did similar calculations based on magnetic resonance imaging (MRI). Differentiation between the formant frequencies and resonance frequencies of the vocal tract can be found in some papers comparing measurements from phonation (formants) to those derived from vocal tract impedance measurements or from calculations based on MRI or computer tomography (CT) data (resonance frequencies) (e.g., Stoffers *et al.*, 2006; Vampola *et al.*, 2013).

What is relevant here for nomenclature and symbolic notation is that the letter R is easily distinguishable from the letter F or f, both in speaking and writing. Hence, it is useful as a subscript to separate source and filter symbols. Discussion can continue on whether or not a formant is a meaningful representation of any particular resonance. Some authors describe resonances pertaining to the supraglottal airway only (assuming no coupling to the glottal or subglottal system), while others describe the net effect of complex interactions of multiple resonators above, below, and within the larynx.

II. IDEAL CLARITY FOR MATHEMATICAL COMMUNICATION

Assuming that a common definition between formants and resonances of some portion of the vocal tract could be reached, the formant frequencies could continue to be written as $F_1, \ldots, F_m$, and source frequencies could be written as $f_1, f_2, f_3, \ldots, f_n$. Source amplitudes could be written as A_n, and relative levels of formant peaks in the vocal-tract spectral envelope as $L_1, \ldots, L_n$. This kind of simplicity is a hope of many scientists and practitioners in the field. Unfortunately, the common definition between a formant and a resonance is yet to be established. Furthermore, for oral communication, there is no upper and lower case distinction. "f_1" is pronounced the same way as "F_1." So, what is the answer for current scientific and oral clarity?

III. A REASONABLE COMPROMISE FOR WRITTEN AND SPOKEN COMMUNICATION

The present authors suggest the following notation to be used. Harmonic frequencies should be written as multiples of f_o, namely, nf_o. The letters H and h are not needed. Harmonic amplitudes should be written as A_n. The letter R can be used as a word abbreviation for resonance, but two subscripts should be assigned to specify the resonance properties (see Table I). The letter F can be used as a word abbreviation for formant, but if only a single subscript is assigned, it must refer only to formant frequency (Table I). Level and bandwidth of the formant should carry two subscripts to be distinguishable from those of resonances.

The harmonic notation is tied to the Fourier series expansion of an acoustic pressure

$$P(t) = A_n \sin(2\pi n f_0 + \phi_n). \quad (1)$$

The parenthesized (1) for the first harmonic in Table I is generally not written or spoken, but always implied. This is important to point out so that the harmonic integer series is complete. The subscript for fundamental frequencies is an "o," not a zero to emphasize "oscillation." The letter L for resonance level is used because we usually express relative formant peaks in dB. L_1 is generally assumed to be 0 dB, thereby using the amplitude of the fundamental as the reference amplitude.

For inharmonic source frequencies, the symbols $f_1, f_2, f_3, \ldots$, can be used without reference to any harmonic index or f_o. It is then important to speak "source frequency f_1," "source frequency f_2," etc. For subharmonic frequencies, nf_o/i will identify the period-i subharmonic series. Consensus on symbols for amplitudes and levels of subharmonics has not yet been discussed.

A little training will be needed for people to say, "two f_o," "three f_o," and so on, for harmonic frequencies. The beauty of that training, however, is that the harmonic relationship with f_o will always be kept in mind. Also, for subharmonics, "one-half f_o" or "one-third f_o" is an easy extension. Speaking the extra letters in $f_{R1}, \ldots, f_{Rm}$; $L_{R1}, \ldots, L_{Rm}$; and $B_{R1}, \ldots, B_{Rm}$ will also be an immediate reminder of "resonance" rather than "source" or "formant." With this nomenclature, an nf_o/f_{Rm} ratio or an $nf_o - f_{Rm}$ difference describes a source-resonance frequency relation. Likewise, an nf_o/F_m or an $nf_o - F_m$ describes a source-formant frequency relation. The ratio A_n/A_1 describes a harmonic relation (linear scale), $L_n - L_1$ describes a logarithmic (dB) source-harmonic relation, $L_n - L_{Rm}$ describes a logarithmic source-resonance amplitude relation in dB, and $L_{Fm} - L_{F1}$ describes a formant level relation in dB.

The classical equation for the resonance frequency of an idealized, uniform closed-open tube

$$f_{Rm} = (2m - 1)\frac{c}{4L} \quad (2)$$

remains a benchmark of comparison between resonances and formants, inasmuch as no resonance coupling occurs to other airway structures. For this idealized airway structure, as well as for closed glottis vowels, $f_{Rm} = F_m$.

Symbolic notation for subglottal resonances has not been addressed here, nor for resonances of side branches of the airways (nasal tract, sinuses). Some precedence exists for labeling subglottal resonance and formant frequencies with a "prime" superscript (e.g., f'_{Rn} and F'_n). Subscripts "sg" have

TABLE I. Harmonic, resonance, and formant symbols for quantitative relations.

Harmonics			Resonances			Formants		
Frequency (Hz)	Amplitude (Pa)	Level (dB)	Frequency (Hz)	Level of peak (dB)	Bandwidth (Hz)	Frequency (Hz)	Level of peak (dB)	Bandwidth (Hz)
$(1)f_o$	A_1	L_1	f_{R1}	L_{R1}	B_{R1}	F_1	L_{F1}	B_{F1}
$2f_o$	A_2	L_2	f_{R2}	L_{R2}	B_{R2}	F_2	L_{F2}	B_{F2}
$3f_o$	A_3	L_3	f_{R3}	L_{R3}	B_{R3}	F_3	L_{F3}	B_{F3}
...								
nf_o	A_n	L_n	f_{Rm}	L_{Rm}	B_{Rm}	F_m	L_{Fm}	B_{Fm}

also been used in presentations, but these additional subscripts are unappealing due to overuse of subscripts.

IV. CONCLUSION

A compromise has been reached between preserving historical nomenclature and symbols for source harmonics, vocal tract resonances, and formants while providing clarity for speaking the symbols and assigning numbers and units to them. Little re-training is needed. One extra subscript is added for resonance characteristics and for formant levels and bandwidths. The harmonic number is explicitly written and spoken together with the fundamental frequency. Authors who are heavily invested in formant frequency analysis are encouraged to be as clear as possible about the relation between a peak in the output spectrum and a presumed resonance of the vocal tract. Likewise, those who describe airway resonances are encouraged to be as clear as possible about their manifestation in the output spectrum. It is important to clarify what the boundaries of the resonator are. In some cases, only the supraglottal vocal tract is described as a resonator (with the glottis closed), in other cases the resonance includes the interaction with the glottis, and in yet other cases resonance includes the entire airway, lungs to lips. As benchmarks are being developed for characteristic frequencies and bandwidths of vowels and consonants, across species, genders, age and cultures, it becomes ever more important to define the exact geometry and boundary conditions of the portion of the airway under investigation.

ACKNOWLEDGMENT

Support for this project was provided by NIH NIDCD Grant No. 1 R01 DC012045-01.

ANSI (**2004**). ANSI S1.1-1994, *American National Standard Acoustical Terminology* (Acoustical Society of America, Melville, NY).

Badin, P., and Fant, G. (**1984**). "Notes on vocal tract computation," *Quarterly Progress and Status Report, Department for Speech, Music, and Hearing*, KTH, Stockholm, pp. 53–108.

Benade, A. H. (**1976**). *Fundamentals of Musical Acoustics* (Oxford University Press, New York), pp. 430–501.

Cohen, E. R., and Giacomo, P. (**1987**). "Symbols, units, nomenclature and fundamental constants in physics," document I.U.P.A.P.-25 (SUNAMCO 87-1) (International Union of Pure and Applied Physics, SUNAMCO Commission, London, UK), available at http://iupap.org/wp-content/uploads/2014/05/A4.pdf (Last viewed 02/10/2015).

Fant, G. (**1960**). *The Acoustic Theory of Speech Production* (Moulton, the Hague), p. 20.

Flanagan, J. L. (**1965**). *Speech Analysis, Synthesis and Perception* (Academic, New York), p. 57.

Hermann, L. (**1894**). "Nachtrag zur Untersuchung der Vocalcurven ("Addendum to the investigation of vowel lines")," Arch. Ges. Physiol. **58**, 264–279.

Hermann, L. (**1895**). "Weitere Untersuchungen über das Wesen der Vocale ("Further analysis about the characteristics of vowels")," Pflügers Arch. Eur. J. Physiol. **61**(4), 169–204.

Russell, G. O. (**1929**). "Mechanism of speech," J. Acoust. Soc. Am. **1**(1), 32–33.

Stevens, K. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA), p. 131.

Stoffers, J., Neuschaefer-Rube, C., and Kob, M. (**2006**). "Comparison of vocal tract resonance characteristics using LPC and impedance measurements," Acta Acust. Acust. **92**(5), 689–699.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. of Am. **100**(1), 537–554.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice Hall, Englewood Cliffs, NJ) [2nd printing (2000). (National Center for Voice and Speech, Salt Lake City, UT)], p. 156.

Vampola, T., Horacek, J., Laukkanen, A.-M., and Švec, J. G. (**2013**). "Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modelling based on CT measurement," Logoped. Phoniatr. Vocol. **40**(1), 14–23.

Wolfe, J. (**2014**). www.phys.unsw.edu.au/jw/formant.html (Last viewed 01/05/2015).