JASA ARTICLE

Lip hyper-articulation in loud voice: Effect on resonance-harmonic proximity

Maëva Garnier,[1,a)] John Smith,[2] and Joe Wolfe[2]
[1]Univ. Grenoble Alpes, CNRS, Grenoble Institute of Engineering Univ. Grenoble Alpes, GIPSA-Lab, 38000 Grenoble, France
[2]School of Physics, UNSW Sydney, Sydney, New South Wales 2052, Australia

ABSTRACT:
Men and women speakers were recorded while producing sustained vowels at comfortable and loud levels. Following comfortable speech, loud levels were produced in three different conditions: first without specific instruction (UL); then maintaining the same pitch as the comfortable level (PL); and finally, keeping both pitch and lip articulation constant (PAL). The sound pressure level, the fundamental frequency (f_o), the first two vocal tract resonances (R1 and R2), the lip geometry, and the larynx height were measured. For women, a closer proximity of R1 to its nearest harmonic, nf_o, was observed in UL. However, no such increased proximity was found in PL, when speakers could, and did, hyper-articulate. Also, no increased proximity was observed in PAL, when lip articulation was constrained. No significant increase in R1:nf_o proximity was observed in men in any of the three loud conditions. Finally, R2 was not observed significantly closer to a voice harmonic in loud speech, for neither men nor women. https://doi.org/10.1121/10.0016595

(Received 11 July 2022; revised 29 November 2022; accepted 1 December 2022; published online 19 December 2022)
[Editor: Zhaoyan Zhang] Pages: 3695–3705

I. INTRODUCTION

Loud or shouted speech involves significant modifications of speech production at aerodynamic, glottal, and articulatory levels: First, speakers increase subglottal pressure, which not only has direct consequences on the level of the output sound (Isshiki, 1964; Plant and Younger, 2000), but also contributes, along with other accompanying or compensatory glottal modifications (Holmberg *et al.*, 1988; Sundberg, 1987), to an increase in the voice's fundamental frequency (f_o) (Gramming *et al.*, 1988; Plant and Younger, 2000; Titze, 1989) and a voice spectrum with a flattened spectral tilt (Raitio *et al.*, 2013; Rostolland, 1982a; Sundberg and Nordenberg, 2006).

Concomitant modifications of the vocal tract articulation are also observed with increased vocal effort. In particular, vowels in loud speech are articulated with an increased mouth aperture and a higher vertical position of the larynx (Garnier *et al.*, 2018; Geumann, 2001; Schulman, 1989), both associated with an increase in the frequency of the first vocal tract resonance (R1) and of the formant (F1) it produces (Bond and Moore, 1990; Lienard and Di Benedetto, 1999; Rostolland, 1982b; Traunmüller and Eriksson, 2000). Variation of the second formant frequency (F2) is less consistent and depends on languages and vowel categories. The simultaneous changes in f_o and in the resonances may alter the relative distribution of resonances and harmonics. The present study tests the hypothesis that lip hyper-articulation and raised f_o of loud speech, commonly observed in any speaker, contribute to an increased proximity between vocal tract resonances and voice harmonics, which may facilitate the production of loud sounds.

This hypothesis originates from the previous observation of comparable lip hyper-articulation in singers when they vocalize in their high range and/or at loud intensity. This hyper-articulation was consistent with formant tuning, or resonance tuning: a systematic proximity of vocal tract resonances (usually R1, sometimes R2) to voice harmonics (often f_o, sometimes higher harmonics nf_o). For example, R1:f_o tuning has been observed in a professional woman singer (Sundberg, 1975) but has been subsequently observed in trained and untrained singers (Joliveau *et al.*, 2004; Garnier *et al.*, 2010; Henrich *et al.*, 2011). R1:nf_o tuning for $n > 1$ has been reported in both men and women singers of different musical styles (Bourne *et al.*, 2016; Bourne and Garnier, 2012; Henrich *et al.*, 2007; Henrich *et al.*, 2011; Sundberg *et al.*, 2011). Resonance tuning has also been suggested in the case of some stage actors (Raphael and Scherer, 1987). Such a closer proximity between vocal tract resonances and voice harmonics is expected to enable the production of loud phonation with less effort. First, the proximity would boost the amplitude of the tuned harmonic and its associated formant and, therefore, contribute to the increased overall intensity (Titze and Sundberg, 1992; Vurma, 2022). Second, it could optimize the energy exchange between the glottal source and the vocal tract filter and stabilize vocal fold vibration (Titze, 2001, 2004, 2008). The supra-glottal acoustic load at the frequency of a voice harmonic nf_o is inertive (or mass-like with acoustic pressure leading acoustic flow) when that harmonic lies close to but below the nearest vocal tract resonance Ri, whereas the load

[a)]Electronic mail: maeva.garnier@gipsa-lab.grenoble-inp.fr

https://doi.org/10.1121/10.0016595 JASA

is compliant (spring-like, with pressure lagging flow) when it is close to but above nf_o. It is, thus, interesting to ask whether any enhanced resonance-harmonic proximity could also be observed in speech production at loud intensity and by speakers without any specific vocal training.

Variations of voice intensity, frequency, and articulation are inter-related. However, since all three variables can be consciously controlled, this allows, to some extent, the possibility of measuring some of the relations with one or more variables held constant. Therefore, three different experimental conditions for the production of loud speech were studied. In the first condition, changes in lip articulation and resonance-harmonic proximity were measured when participants went from a comfortable to a loud voice in "natural" or unconstrained phonation [condition "unconstrained loud" (UL)]. The second required participants to keep the pitch constant when vocal intensity was increased [condition "pitch-constrained loud" (PL)]; this should enable the examination of how articulatory modifications of loud speech might, on their own, contribute to any increase in resonance-harmonic proximity. For the third condition, vocal intensity was increased with both pitch and lip articulation constrained [pitch and lip articulation-constrained loud (PAL)]; in the case of an increased resonance proximity in the previous conditions UL and PL, this last condition PAL should enable us to verify that this proximity is controlled by variations in lip articulation: If this is the case, no increase in resonance-harmonic proximity is expected in that condition, when lip articulation is constrained.

II. MATERIALS AND METHODS

A. Participants

In a study approved by the Human Ethics Committee of UNSW Sydney, eight native speakers of Australian English, four men (M1, M2, M3, M4) and four women (W1, W2, W3, W4), aged 22–65 years old, participated in the study without payment. None of them reported any hearing or voice disorders. None reported any academic knowledge about voice production, hyper-articulation, or formant tuning phenomena. They were only informed that they would undertake an experiment in speech production.

B. Experimental protocol

The corpus consisted of six Australian vowels: [æ], [ɛ], [u], [ʊ], [ɔ], and [ɜ], described to the participants as the vowels in the words /had/, /head/, /hoot/, /hood/, /horde/, and /heard/, respectively (no further description or context for the vowels was given). [æ] and [ʊ] were chosen as extremes of the vowel system. The third extreme vowel [i] was not selected here, because R1 for that vowel is less easy to measure accurately with vocal tract spectrometry (Epps *et al.*, 1997). The vowels [ɛ], [u], and [ɔ] were selected as intermediate vowels between [æ], [i], and [ʊ] in the (F2, F1) plane. The vowel [ɜ] was considered as a central vowel in the Australian English vowel system. Each vowel was first sustained for 5 s at a comfortable level (Comf) and then, in the same breath, for an additional 5 s at loud level (as if addressing a person situated 20 m ahead). An audio beep indicated to the participants when to start increasing vocal effort and when to stop phonation (see Fig. 1).

Three conditions were recorded.

- In a first session (UL), participants were told to increase vocal intensity after the beep.
- In a second session (PL), participants were asked to keep a constant pitch when increasing vocal sound pressure level (SPL). Such a task was not easy for most of the speakers, who did not have any particular vocal expertise, so that the experimenter gave them some examples, and the participants were asked to practice for approximately 5 min before being able to record the task.
- In a third session (PAL), participants were instructed to increase vocal SPL while keeping both pitch and lip articulation constant. Again, a brief practice session was necessary before recording that condition.

Each vowel was produced five times in every session, at both comfortable and loud level, before moving on to the next session. The order of the three conditions was the same for all the participants, so they were unaware of the constraint(s) to be imposed in the next condition. To prevent the development of vocal fatigue (the whole protocol lasted for about 45 min, including about 25 min of phonation), participants were instructed to take regular breaks and were provided with water throughout the recording session to keep their mouths and throats hydrated.

C. Measurements

Five signals were simultaneously recorded:

- The audio signals were recorded with two 1/4-in. pressure microphones (4944-A, Bruël and Kjær, Nærum, Denmark), one located at the participant's lips, the second placed 30 cm in front of him/her. Both signals were then

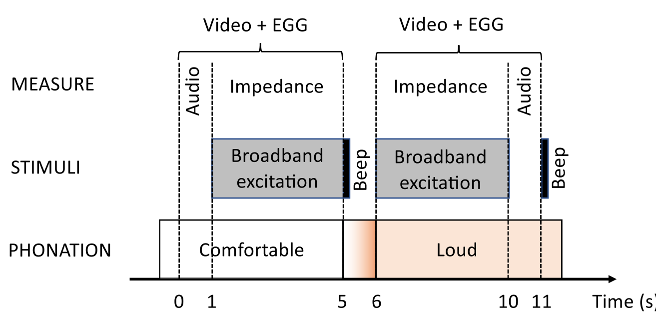


FIG. 1. (Color online) Illustration of the protocol. Participants produced sustained vowels for about 11 s with an increase in vocal SPL after 5 s in a same breath. Audio beeps (black bars) indicated to the participants when to start loud voice and when to stop phonation. The vocal tract was excited with a calibrated broadband sound signal during 4 s of comfortable and loud phonation to measure the vocal tract impedance spectrum. Voice SPL and spectrum were characterized from the audio signal, during the remaining second of "clean" phonation at comfortable and loud levels. The laryngeal behavior and lip articulation were characterized over the whole 5 s interval of comfortable or loud phonation, based on electroglottographic measurements and video images.

amplified (Bruël and Kjær Nexus 2690) and digitized at 16 bits and a rate of 44.1 kHz using a FireWire audio interface (MOTU 828, Cambridge, MA). The lip microphone was attached alongside a slightly flexible tube, 8 mm inner diameter, that was connected to a loudspeaker via an impedance matching horn. This acoustic source was used to excite the vocal tract with a synthesized broadband signal for the last 4 s of phonation at Comf and the first 4 s of phonation at loud level (see Fig. 1). The broadband signal was synthesized as the sum of sine waves over the range of 200–3000 Hz spaced at 10.8 Hz ($= 44.1$ kHz/2^{12}) with amplitudes adjusted to improve the signal-to-noise ratio and relative phases adjusted to improve dynamic range (Joliveau *et al.*, 2004; Smith, 1995). During the 4 s when the acoustic source was used, the microphone recorded the response of the vocal tract to that excitation to measure the vocal tract impedance and to detect from it the frequency of vocal tract resonances [see Epps *et al.* (1997) and Garnier *et al.* (2010) for details on vocal tract impedance spectrometry]. A stand was adjusted for height so that the microphone and the flexible tube to which it was attached rested gently against the participant's lower lip during measurements. The microphone situated 30 cm away from the lips was also attached to a stand adjusted to the height of the participant's mouth. Its signal was used to measure vocal SPL and spectrum descriptors during the "clean" second of phonation, when there was no broadband excitation (see Fig. 1).

- The electroglottographic (EGG) signal and the larynx tracking (LT) signals were recorded with an electroglottograph EG2 (Glottal Enterprises, Syracuse, NY), using medical conductive gel to improve electrical contact between the skin and the electrodes. Two pairs of electrodes were placed, one on each side of the thyroid cartilage at the height of the cartilage while the participants were phonating in their comfortable middle range. No automatic gain control was used. The high-pass filter was set to a 10 Hz cutoff frequency. Both signals were then digitized at 16 bits and a rate of 44.1 kHz using the same audio interface as for the audio signal. Since the LT signal presents slow variations, below the cutoff frequency of the audio interface, it was acquired by modulating the amplitude of a 10 kHz carrier wave.
- Front images of the speaker's lips were recorded at a rate of 25 images/s, using a standard video camera mounted on a stand and placed at a 30 cm distance from the speaker's face and at the same height. The camera's audio signal, compared by cross-correlation with the main audio signal of the microphone at 30 cm, enabled post-synchronization of the articulatory measurements with other audio and EGG data.

1. SPL

The mean SPL was measured precisely from the clean second of comfortable and loud phonation (no broadband excitation; see Fig. 1), using the internal calibration signal of 1 V_{RMS} at 1 kHz delivered by the conditioning amplifier and knowing its V/Pa transduction coefficient. It is expressed in dB(Z) with respect to 20 μPa, with (Z) meaning that no weighting was applied to account for the frequency sensitivity of the human ear.

2. f_o

The mean f_o was extracted from the EGG signal over the whole 5 s time interval of comfortable and loud phonation (see Fig. 1), using MATLAB scripts. The EGG signal was first high-pass filtered above 60 Hz and low-pass filtered below 1 kHz. Then the f_o was estimated from the derivative of the electroglottographic signal (DEGG) by detecting glottal closing times (Henrich *et al.*, 2004). Its variations were measured in equal-tempered semitones.

3. Vocal tract resonances

The mean frequency of each of the first two vocal tract resonances (R1 and R2) was measured over a 4 s time interval for both comfortable and loud modes of phonation, while the vocal tract was excited with a broadband signal (see Fig. 1). R1 and R2 were detected visually from local maxima of the measured pressure ratio $\gamma = p_{//}/p_r$, where $p_{//}$ is the pressure spectrum measured with an open mouth (whilst phonating), and p_r is the spectrum measured at the lips with the mouth closed [performed during an earlier calibration measurement for each participant (see Garnier *et al.*, 2010)].

4. Lip geometry

Mean lip spreading (LS; horizontal dimension), lip aperture (LA; vertical dimension) and inter-lip area (ILA) were measured over the same 5 s time interval of the vowel as for glottal measurements (see Fig. 1), using a semi-automatic detection method of the inner edge of the external lip aperture. The distance to the camera remained constant, because the participants kept the microphone and the broadband source on their lower lip. The pixels to millimeters conversion was calibrated prior to the recordings, using a 1 cm $\times$ 1 cm grid.

5. Larynx height

The mean variation in larynx height (ΔLH) was measured from the demodulated LT signal and converted into millimeters, using a calibration method suggested by Pabst and Sundberg (1993).

D. Data analysis

Several statistical analyses were conducted using the R software (R Core Team, 2013). The conventional notation was adopted to report statistical results: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ns, not significant ($p > 0.05$).

For each of the voice descriptors, the variation from comfortable to loud phonation was considered, and an analysis of variance (ANOVA) was conducted from these "Delta" variables, to examine whether the variations of these descriptors were significantly non-null and whether they depend on the VOWEL (six levels: [æ], [ɛ], [u], [ʊ], [ɔ], and [ɜ]) and the SESSION (three levels: UL, PL, and PAL). A group analysis was conducted, based on a mixed model of the data (using the R package lme), considering again two fixed effects (factors VOWEL and SESSION) but also a random effect (factor SPEAKER, on the intercept). The simplest model that best explains the variance of a Delta variable was searched, using a descending approach (function step in R), based on likelihood ratio tests (LRT) and the minimization of the Bayesian information criterion (BIC). Hypotheses about the model's normality and homoscedasticity were validated by looking at the residuals graphs. After examining the effects of the interaction terms remaining in the simplified model, specific contrasts were tested using the multcmp package in R and applying Bonferroni adjustments for multiple comparisons.

For each vowel and gender, the proportion of cases in which R1 or R2 was found closer than ±35 Hz to a voice harmonic was also calculated. An ANOVA was then conducted to examine whether this proportion varies significantly with GENDER and the PHONATION MODE (four levels: Comf and loud voice of UL, PL, and PAL), based on a generalized linear model of the data (using the R package glm). This proportion of cases in our data was also compared to a chance level of observing such resonance-harmonic proximity, calculated from Monte-Carlo simulations for R1 and R2 and both genders.

III. RESULTS

A. Performance of the task

The five graphs of Fig. 2 show the extent to which the speakers complied with the experimental instructions.

As requested, all the participants were able to increase significantly their vocal SPL after the first audio beep of the protocol, for all the vowels and the three sessions (UL, PL, and PAL). However, that increase strongly depended on the session: the increase was much greater during UL (on average $+16.1 \pm 0.6$ dB, $p < 0.001$), when vocal SPL was increased "naturally," compared with PL, when that increase was constrained in pitch only (on average $+8.3 \pm 0.6$ dB, $p < 0.0001$), or PAL with both pitch and lip articulation constrained (on average $+8.7 \pm 0.6$ dB, $p < 0.0001$).

No participant showed a significant variation in f_o from comfortable to loud phonation in the second and third session, for which pitch was supposed to be maintained. The average variation of f_o was of -0.002 ± 0.30 semitones in PL ($p = 1$, ns) and 0.11 ± 0.30 semitones in PAL ($p > 0.9$, ns). In contrast, f_o increased significantly by $+8.2 \pm 0.3$ semitones on average in UL ($p < 0.001$). For men, the average values of f_o increased from 115 to 197 Hz, 128 to 129 Hz, and 128 to 130 Hz for UL, PL, and PAL constrained, respectively. For women, the corresponding changes were from 236 to 362 Hz, 258 to 257 Hz, and 263 to 261 Hz.

All the participants were able to maintain their lip articulation whilst increasing vocal SPL in the third session: LA increased by only $+0.4 \pm 0.2$ mm on average in PAL ($p > 0.09$, ns), and LS increased by $+0.3 \pm 0.5$ mm ($p > 0.7$, ns), whereas it varied significantly in UL [on average, ΔLA $= +5.5 \pm 0.2$ mm ($p < 0.001$) and ΔLS $= +6.3 \pm 0.5$ mm ($p < 0.001$)] and PL [on average, ΔLA $= +2.6 \pm 0.2$ mm ($p < 0.001$) and ΔLS $= +3.4 \pm 0.5$ mm ($p < 0.001$)]. It is interesting to note that, although the requested constraint in PL was only applied to pitch, participants, nevertheless, showed a reduced increase in mouth aperture and spreading: roughly half as much when pitch was constrained as when no constraint was requested.

Participants were not instructed to control the vertical displacements of the larynx. However, none of the participants showed a significant variation in larynx height with the increase in vocal SPL in PAL (on average,

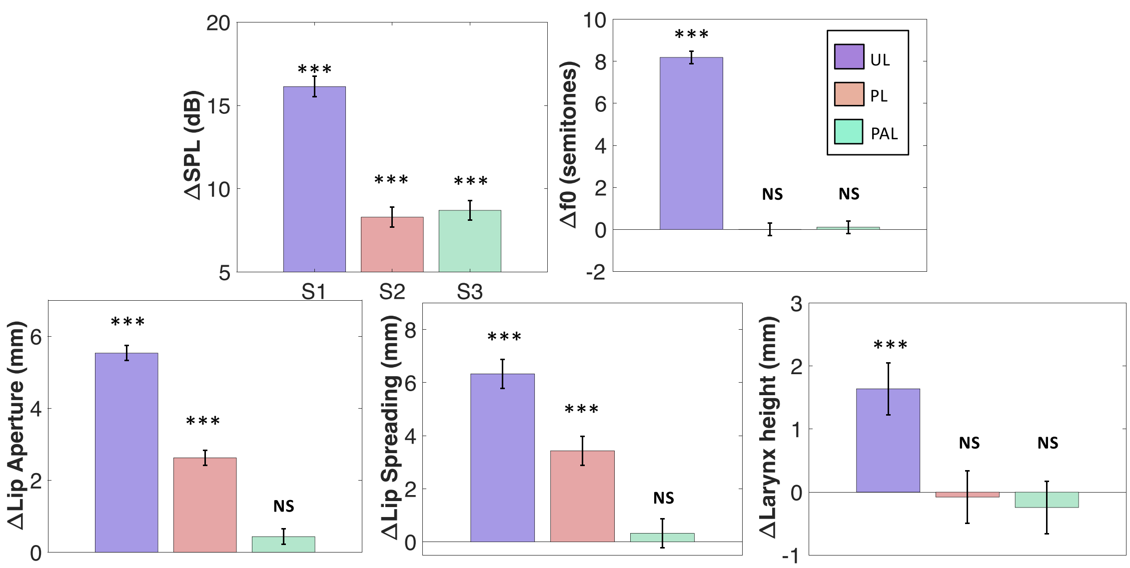


FIG. 2. (Color online) Average change of voice SPL, f_o, Lip Aperture (vertical dimension), Lip Spreading (horizontal dimension), and larynx height with the increase in vocal intensity in the three sessions (UL, PL, and PAL).

–0.2 ± 0.4 mm, $p > 0.9$, ns). In contrast, all the participants showed a significant rise of their larynx in UL (on average, +1.6 ± 0.4 mm, $p < 0.001$). In PL, half of them showed a significant rise, whereas the other half did not, so that no significant variation was observed on average over the whole group (–0.1 ± 0.4 mm, $p > 0.9$, ns).

B. Proximity between R1 and R2 and voice harmonics

The acoustic consequences of R1 and R2 variations on the reorganization of the whole vowel system are shown in Fig. 3. To sum up, the variations of R1 and R2 in naturally loud voice (UL) were in agreement with the variations of F1 and F2 reported in previous studies of loud and shouted speech (Bond and Moore, 1990; Garnier *et al.*, 2018; Lienard and Di Benedetto, 1999; Rostolland, 1982b). They mainly consisted in a shift of the vowel system toward higher frequencies of the first resonance R1 and a tendency toward a slight reduction of the vowel system along the back-front dimension, due to the greater increase in R2 on back rounded vowels.

1. R1

If lip hyper-articulation in loud speech had the effect of "tuning" the first resonance frequency R1 close to a voice harmonic, the distribution of R1 frequencies would show unusually many cases where $Ri \sim nf_o$ in UL and PL, when participants could adjust their lip articulation, compared with Comf, and when also compared with PAL, for which the participants maintained lip articulation constant.

As expected, all the participants showed a significant increase in R1 when increasing vocal SPL in UL (+22.8 ± 2.4% on average, $p < 0.001$) (see Fig. 3). There was no significant difference between men and women (Δ = –4.6 ± 5.2%, $p = 0.41$). This percentage increase did not depend significantly on the vowel [LRT: degrees of freedom (df) = 5, LRatio = 12.95, $p = 0.23$]. A significant increase in R1 was also observed in PL (+9.3 ± 2.4%, $p < 0.001$) and PAL (+9.1 ± 2.4%, $p < 0.001$) when pitch was maintained constant, although it was considerably reduced compared with UL. A surprising result was that the increase in R1 was still significant in PAL and comparable with that observed in PL, although measured lip articulation was maintained in that PAL session and larynx height did not vary significantly (see Fig. 2).

a. For women. Figure 4 shows, for women, the distribution of the frequency difference $R1 - nf_o$, where R1 is the frequency of R1, and nf_o is that of the closest harmonic. The shaded area corresponds to a resonance bandwidth of ~70 Hz (Hanna *et al.*, 2016), where a voice harmonic is expected to benefit substantially from a boosting effect of the resonance. If the vocal tract resonances systematically fell close to voice harmonics (which we call resonance proximity), one would observe a single peak distribution with a mode situated in this shaded central area.

For women, distributions of the frequency difference $R1 - nf_o$ in comfortable speech do not demonstrate a monomodal shape with a high central peak (see Fig. 4), meaning that R1 is not systematically close to a voice harmonic. Figure 4 instead shows that, in comfortable speech and the context of this experiment, the R1 for women do not often lie within the bandwidth of the $R1 = nf_o$ lines. During UL, however, a centered and single peak distribution of the $R1 - nf_o$ frequency difference is observed for the mid-open vowels [ɛ], [ɜ], and [ɔ]. Figure 5 shows in more detail that these cases of resonance proximity involve R1 and the second voice harmonic ($2 f_o$), which

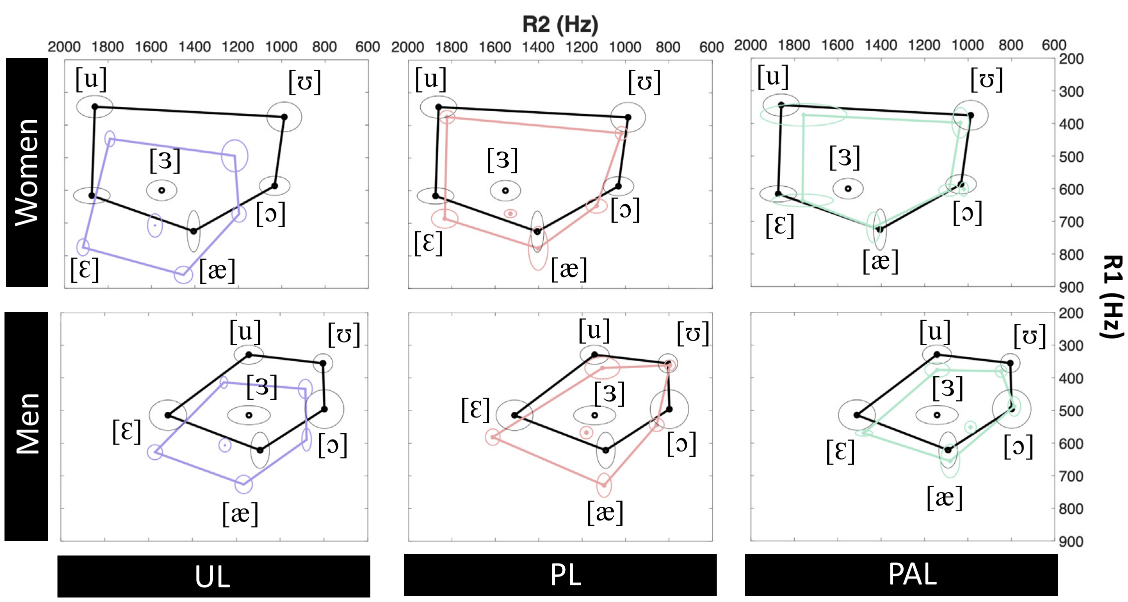
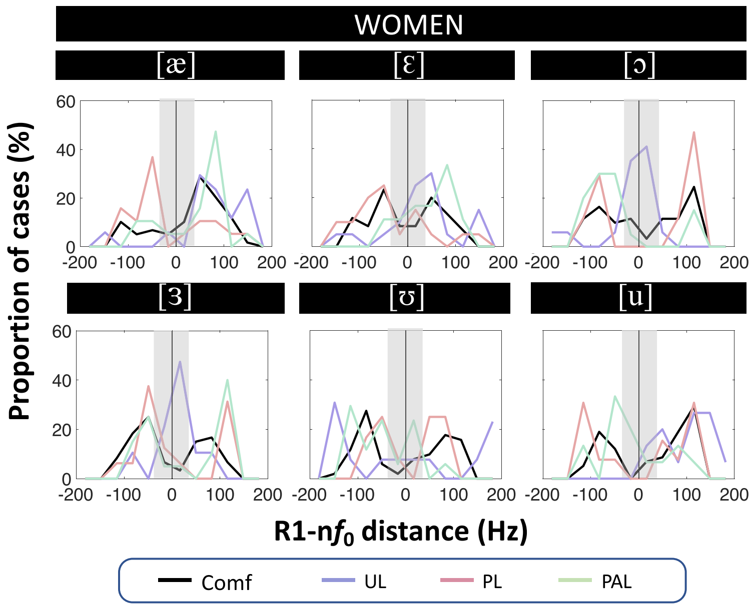
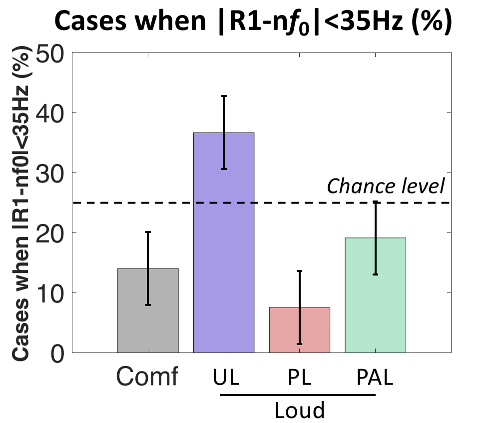


FIG. 3. (Color online) Average frequency of the first two vocal tract resonances (R1, R2) showing the effect of increasing vocal effort from comfortable speech (black) to loud speech (light color) in the three sessions (UL, PL, and PAL) for all the six vowels and both genders. The ellipses represent the average intra-speaker variability within each vowel category.

JASA





	[æ]	[ɛ]	[ɔ]	[ɜ]	[ʊ]	[u]
Comf	16,4%	20,0%	18,0%	13,3%	8,2%	8,3%
UL	5,6%	45,0%	65,0%	68,4%	21,1%	15,0%
PL	4,8%	20,0%	0,0%	15,0%	5,6%	0,0%
PAL	14,3%	27,8%	5,0%	15,0%	27,8%	25,0%

FIG. 4. (Color online) Left: The distribution of frequency differences observed between the frequency of R1 and the closest voice harmonic (nf_o) for women pronouncing each vowel and mode of phonation [comfortable voice and loud phonations (UL, PL, PAL)]. Right: Percentage of cases for which R1 frequencies are closer than 35 Hz to a voice harmonic, on average, and for each vowel produced by women at comfortable and loud levels. The dashed line represents the level of chance for observing such a resonance-harmonic proximity.

occurs for some speakers and some vowels. For mid-open vowels, the centered and single peak distributions, evident for UL, are no longer apparent in PL and PAL, meaning that no significant resonance proximity is observed in these constrained modes of loud phonation. The situation is different for the close vowels [u] and [ʊ], where the increase in (f_o, R1) only slightly increases the average proximity to the first and second harmonics in UL. For the open vowel [æ], the increase in (f_o, R1) shifts the average away from the higher harmonics in UL.

These qualitative observations are corroborated by further analysis of the proportion of cases for which $|R1 - nf_o| < 35$ Hz and its variation with vowel and phonation mode (see Fig. 4). On average over the different vowels, R1 was found close to a voice harmonic significantly more frequently in the "natural" loud voice (UL), compared with comfortable speech ($\Delta_{UL\,-\,Comf} = 22.6 \pm 8.6\%$, $p = 0.041$), and with a proportion of cases in UL above chance level (25%). No significant variation in the proportion of cases was observed in the two other constrained modes of loud

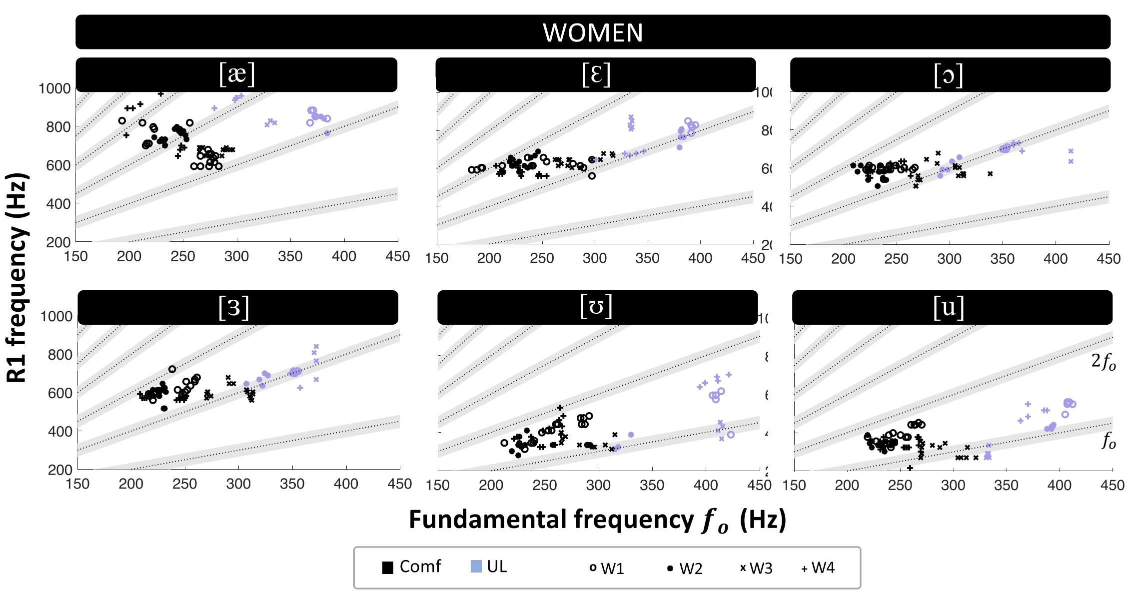
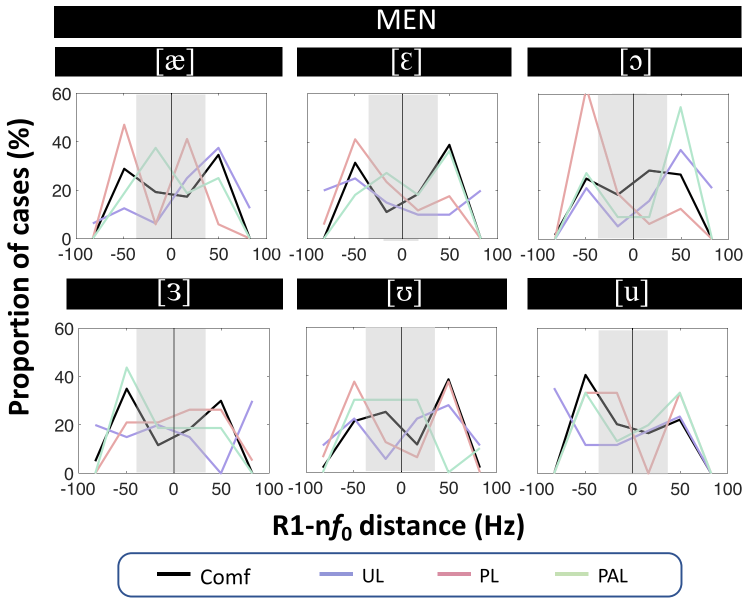
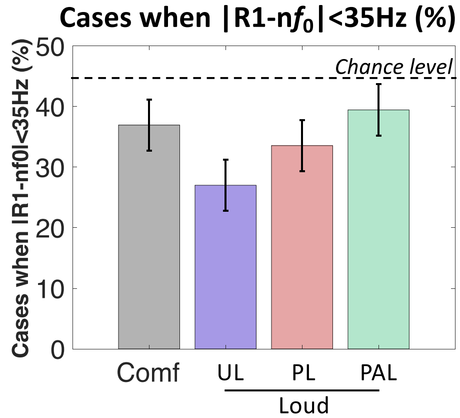


FIG. 5. (Color online) The relationship between values of (f_o, R1) measured simultaneously for the six vowels produced by the four women of this study (W1, W2, W3, W4), in comfortable voice (black) and natural loud voice UL (light blue). The dashed lines represent the voice harmonics nf_o. Shaded gray areas indicate a 70 Hz interval centered around these voice harmonics, corresponding to the typical bandwidth of vocal tract resonances (Hanna *et al.*, 2016).

JASA
https://doi.org/10.1121/10.0016595





	[æ]	[ɛ]	[ɔ]	[ɜ]	[ʊ]	[u]
Comf	37,1%	30,6%	51,7%	30,0%	36,7%	35,6%
UL	22,7%	25,0%	20,0%	40,0%	25,0%	29,4%
PL	36,4%	47,4%	20,0%	45,0%	26,3%	26,3%
PAL	42,1%	60,0%	16,7%	35,3%	38,9%	43,8%

FIG. 6. (Color online) Left: Average distributions in men, for each vowel and mode of phonation [comfortable voice and loud phonation (UL, PL, PAL)], of the frequency differences observed between the frequency of R1 and the closest voice harmonic (nf_o). Right: Percentage of cases for which R1 frequencies are closer than 35 Hz to a voice harmonic, on average, and for each vowel produced by men at comfortable and loud levels. The dashed line represents the level of chance for observing such a resonance-harmonic proximity.

voice ($\Delta_{PL\ -\ Comf} = -6.5 \pm 8.6\%$, $p = 0.79$; $\Delta_{PAL\ -\ Comf} = 5.1 \pm 8.6\%$, $p = 0.88$), with a proportion of cases that was not greater than chance level.

b. For men. $|\mathrm{R1} - nf_o|$ separations for men were globally smaller than for women because their lower f_o gave a reduced separation between voice harmonics. No monomodal, centered distributions of the separations $\mathrm{R1} - nf_o$ were observed in men for any vowel produced at Comf (see Fig. 6). Further, no significantly increased resonance proximity was observed in natural loud voice (UL) or in constrained loud phonation (PL and PAL). The proportion of cases observed was also never greater than chance level (47%). Figure 7 shows with more detail how R1 frequencies varied from Comf to UL without "following" any particular voice harmonic.

Further statistical analysis confirmed that on average over all the vowels, R1 for men was not found close to a voice harmonic significantly more frequently in any of the loud modes of phonation, compared with comfortable speech ($\Delta_{UL\ -\ Comf} = -9.9 \pm 6.0\%$, $p = 0.26$; $\Delta_{PL\ -\ Comf} = -3.4 \pm 6.0\%$, $p = 0.90$; $\Delta_{PAL\ -\ Comf} = 2.5 \pm 6.0\%$, $p = 0.95$) (see Fig. 6).

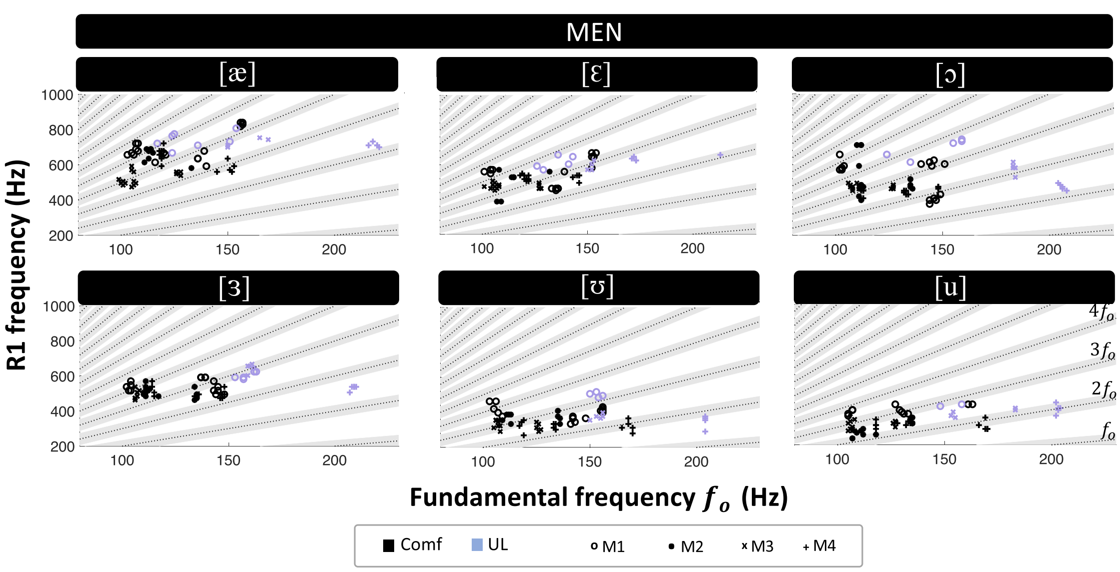
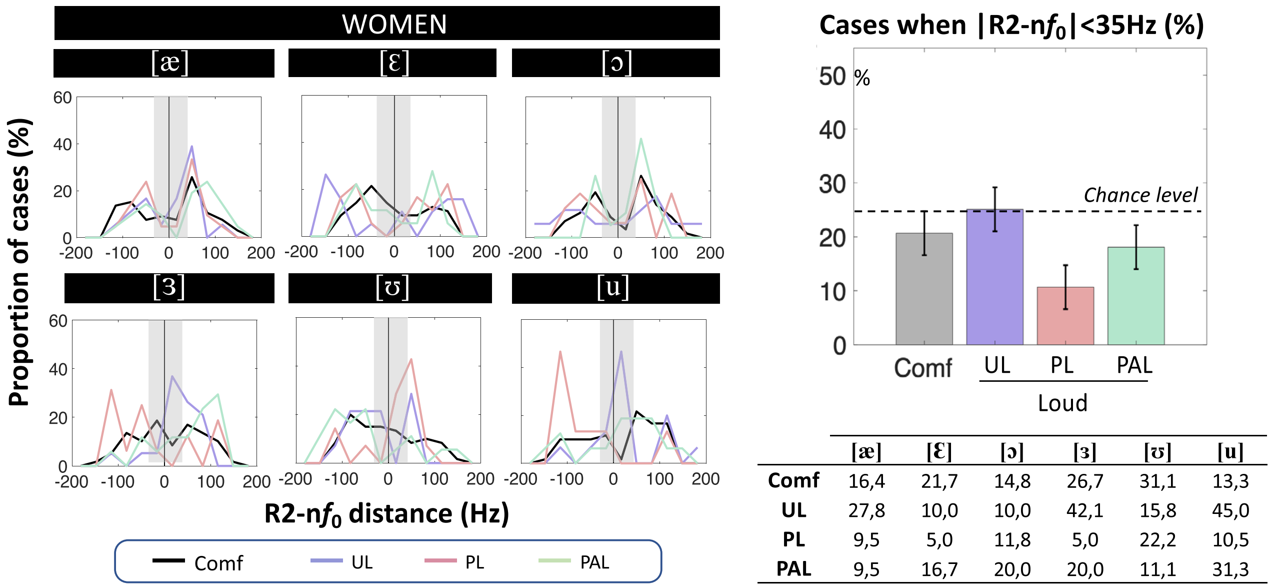


FIG. 7. (Color online) The relationship between values of (f_o, R1) measured simultaneously for the six vowels produced by the four men of this study (M1, M2, M3, M4), in comfortable voice (black) and natural loud voice UL (light blue). The dashed lines represent the voice harmonics nf_o. Shaded gray areas indicate a 70 Hz interval centered around these voice harmonics, corresponding to the typical bandwidth of vocal tract resonances (Hanna *et al.*, 2016).

JASA



	[æ]	[ɛ]	[ɔ]	[ɜ]	[ʊ]	[u]
Comf	16,4	21,7	14,8	26,7	31,1	13,3
UL	27,8	10,0	10,0	42,1	15,8	45,0
PL	9,5	5,0	11,8	5,0	22,2	10,5
PAL	9,5	16,7	20,0	20,0	11,1	31,3

FIG. 8. (Color online) Left: Average distributions in women, for each vowel and mode of phonation [comfortable voice and loud phonation (UL, PL, PAL)], of the difference observed between the frequency of R2 and nf_o. Right: Percentage of measurements for which R2 frequencies are closer than 35 Hz to a voice harmonic, on average, and for each vowel produced by women at comfortable and loud levels. The dashed line represents the level of chance for observing such a resonance-harmonic proximity.

2. R2

a. For women. For women, distributions of the frequency difference R2 $- nf_o$ in comfortable speech did not demonstrate a monomodal shape with a high central peak (see Fig. 8), meaning that R2 frequencies are not systematically tuned close to a voice harmonic. Figure 9 indeed shows that, for women, harmonics in comfortable speech do not often lie within a typical resonance bandwidth of the R2 $= nf_o$ lines. During UL, however, a centered and single peak distribution of the R2 $- nf_o$ frequency difference is observed for the vowels [ɜ] and [u]. Figure 9 shows in more detail that these cases of resonance proximity corresponded to a proximity of R2 to the fourth, fifth, and sixth harmonics, depending on the speaker and the vowel. Finally, Fig. 8 does-not show centered and single peak distributions of the R2 $- nf_o$ frequency difference in PL and PAL, meaning that no significant resonance proximity was observed any longer in these constrained modes of loud phonation.

These qualitative observations are corroborated by further analysis of the proportion of cases when $|R2 - nf_o| < 35$ Hz and its variation with vowel and phonation mode (see Fig. 8). This proportion did not vary significantly with

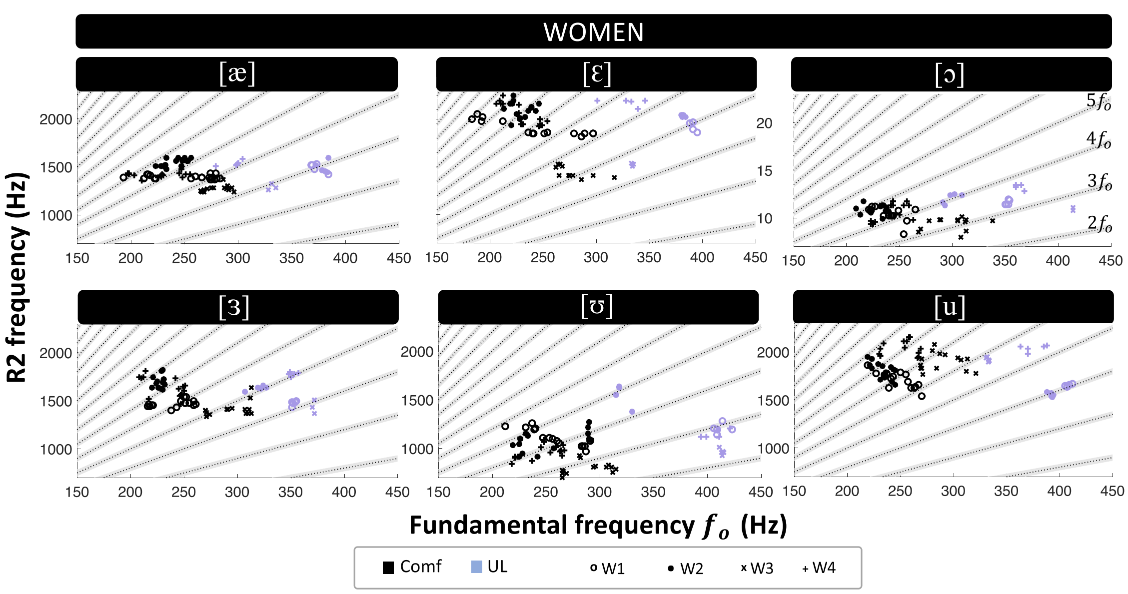
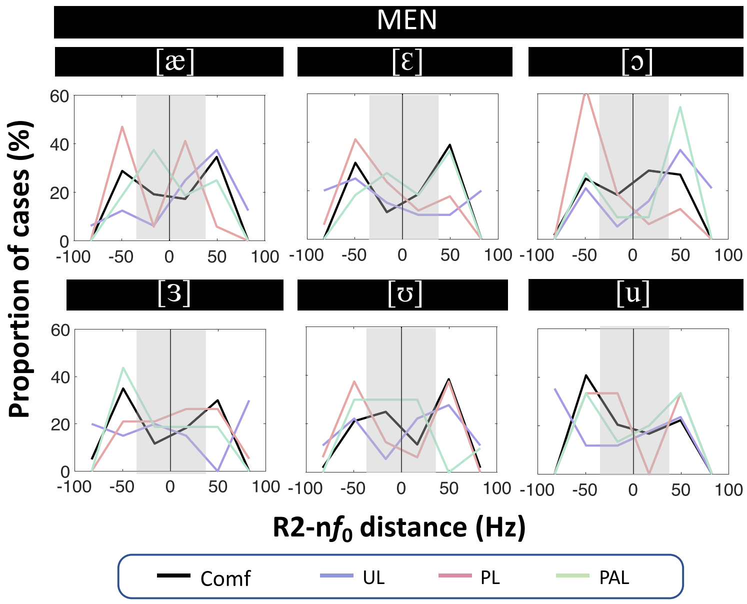
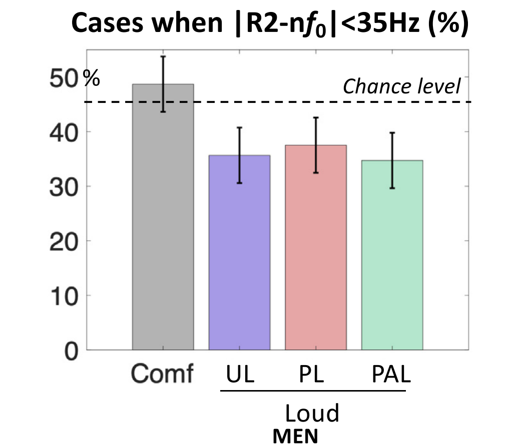


FIG. 9. (Color online) The relationship between values of (f_o, R2) measured simultaneously for the six vowels produced by the four women of this study (W1, W2, W3, W4), in comfortable voice (black) and natural loud voice UL (light blue). The dashed lines represent the voice harmonics nf_o. Shaded gray areas indicate a 70 Hz interval around these voice harmonics, corresponding to the typical bandwidth of vocal tract resonances (Hanna *et al.*, 2016).

https://doi.org/10.1121/10.0016595





MEN

	[æ]	[ɛ]	[ɔ]	[ɜ]	[ʊ]	[u]
Comf	51,4	48,4	45,0	51,7	51,7	44,1
UL	13,6	60,0	30,0	35,0	40,0	35,3
PL	18,2	42,1	30,0	40,0	57,9	36,8
PAL	31,6	46,7	16,7	29,4	27,8	56,3

FIG. 10. (Color online) Left: Average distributions in men, for each vowel and mode of phonation [comfortable voice and loud phonation (UL, PL, PAL)], of the difference observed between the frequency of R2 and nf_o. Right: Percentage of cases for which R2 frequencies are closer than 35 Hz to a voice harmonic, on average, and for each vowel produced by men at comfortable and loud levels. The dashed line represents the level of chance for observing such a resonance-harmonic proximity.

the vowel (df = 5, LRatio = 1.11, $p = 0.40$). The statistical analysis showed that for R2, no significant variation in proportion of proximity cases was observed for women in loud voice (UL), compared with comfortable speech, regardless of the session (ΔUL − Comf = +4.4 ± 5.8%, $p = 0.78$; ΔPL − Comf = –10.0 ± 5.8%, $p = 0.23$; ΔPAL − Comf = –2.6 ± 5.8%, $p = 0.94$), and that this proportion of cases was never greater than chance level (24%) anyway.

b. For men. For men, smaller $|R2 - nf_o|$ distances were globally observed than for women, resulting directly from their lower f_o and the consequently reduced distance between voice harmonics. No monomodal, centered, and peaked distribution of the distances R2 − nf_o is observed in men for any vowel produced at Comf (see Fig. 10). No significantly increased resonance proximity is observed in natural loud voice (UL) or in constrained loud phonation (PL and PAL), and the proportion of cases observed was never greater than chance level (46%). Figure 11 shows with more detail how R2 frequencies vary from Comf to UL without following any particular voice harmonic.

Statistical analysis confirmed that for men, R2 was not observed significantly more frequently closer to a voice harmonic in any of the loud modes of phonation, compared

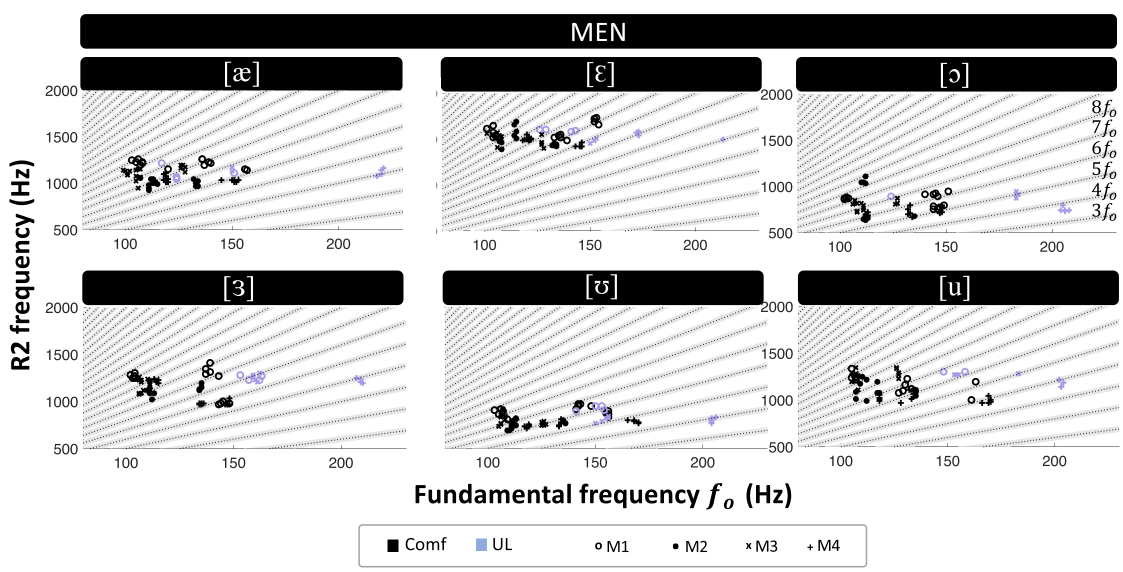


FIG. 11. (Color online) The relationship between values of (f_o, R2) measured simultaneously for the six vowels produced by the four men of this study (M1, M2, M3, M4) in comfortable voice (black) and natural loud voice UL (light blue). The dashed lines represent the voice harmonics nf_o. Shaded gray areas indicate a 70 Hz interval around these voice harmonics, corresponding to the typical bandwidth of vocal tract resonances (Hanna *et al.*, 2016).

with comfortable speech (ΔUL − Comf = +13.0 ± 7.2%, $p = 0.20$; ΔPL − Comf = –11.2 ± 7.2%, $p = 0.30$; ΔPAL − Comf = –14.0 ± 7.2%, $p = 0.16$) (see Fig. 10).

IV. DISCUSSION AND CONCLUSION

When and to what extent are lip hyper-articulation and f_o increase in loud voice associated with increased proximity between a vocal tract resonance and a voice harmonic?

The results of this present study showed, for women only and less than 40% of the time, a significantly more frequent R1:nf_o proximity in the UL condition than in either the comfortable voice or the PAL condition, where speakers increased vocal effort with constrained pitch and lip articulation. However, women did not show a significant increase in R1:nf_o proximity in PL, where loud voice was constrained in pitch but where participants could still adjust their lip articulation. No significant improvement of the R1:nf_o proximity was observed in the loud voice of men, regardless of how vocal intensity was increased [naturally (UL), with constrained pitch (PL), or with constrained pitch and lip articulation (PAL)]. Resonance tuning or resonance proximity would be of limited advantage to men in the low part of their f_o range, where the harmonic spacing f_o is not much smaller than the bandwidth of a vocal tract resonance, so harmonics are more often boosted by a resonance even without adjustment. However, men still showed a significant increase in lip opening in loud voice, accompanied by a significant rise of R1, as did women.

Proximity between R2 and voice harmonics was also examined. For R2, there was no significant adjustment of the resonance close to a voice harmonic when unconstrained loud voice was compared with comfortable voice, for either men or women.

Altogether, these different results suggest that, at least for men, and probably also for women, increasing the resonance-harmonic proximity may not be the main motivation of lip hyper-articulation in loud speech.

This study measured only sustained spoken vowels, so these observations should not be extrapolated to loud speech produced in more normal contexts. Because precise measurements of the resonances require several seconds of phonation, studying resonance-harmonic proximity in conversational speech and ecological situations remains a technical challenge.

A question remains about the observed increase in R1 in the loud voice of PAL, while the lips (and apparently the larynx position; see Fig. 2) were nevertheless maintained as in comfortable speech. This suggests additional contributions to R1 variations from non-visible articulatory modifications, such as oro-nasal coupling or dilation of the pharyngeal cavity, in particular (Xue *et al.*, 2021). Part of this rise could also be due to an increase in average glottal opening (Barney and Stefano, 2007; Swerdlin *et al.*, 2010). Further exploration of tongue, velum, and pharyngeal articulation in loud speech would be needed to better understand these variations of R1.

Another question raised by this work concerns the inter-speaker variability of such "resonance tuning" behaviors in loud voice and its consequence on voice efficiency and laryngeal effort. Indeed, if lip hyper-articulation facilitates the vocal fold vibration in loud phonation—through improved resonance-harmonic proximity (Titze, 2008)—this could result in improved glottal behavior and voice quality when speakers increase vocal intensity with flexible lip articulation, whereas constrained lip articulation might be associated instead with pressed or strained phonation. The results of this study, obtained from eight participants, already show some inter-individual and gender differences in variations of lip aperture, resonance frequencies, and resonance-harmonic proximity with increased vocal intensity. Exploring further this inter-individual variability, not only in healthy individuals, but also in dysphonic patients, may bring some useful information for the prevention and rehabilitation of vocal fatigue and muscle tension dysphonia (Morrison and Rammage, 1993; Roy *et al.*, 2009).

Finally, alternative explanations for the interest or motivations of lip hyper-articulation in loud speech might be envisaged, especially for men for whom no significant improvement of the resonance-harmonic proximity was observed. A first hypothesis is that lip hyper-articulation, and the resulting variation of formants, may enable them to compensate for the concomitant rise of f_o with increased vocal effort, by preserving some relative tonal distances between formant frequencies and f_o, which are considered as perceptual cues to vowel identification and distinction (Hoemeke and Diehl, 1994; Traunmüller, 1981). Not supporting this hypothesis, however, is the fact that lip hyper-articulation, and consequent increase in R1, was still observed here in the PL condition of our experiment, when pitch was constrained and when no compensation for a pitch rise was therefore needed. Another hypothesis is that lip hyper-articulation in loud speech may contribute to improve sound radiation and directivity and, thus, to increase voice intensity as measured in front of the speaker. These different hypotheses deserve to be tested further by dedicated experiments.

ACKNOWLEDGMENTS

We thank our volunteer participants and the Australian Research Council for support, as well as Nathalie Henrich for help with the experimental design, Marine Barbé for help with the analysis of the vocal tract resonances, and Silvain Gerber for his advice on statistical analyses.

Barney, A., and Stefano, A. D. (**2007**). "The effect of glottal opening on the acoustic response of the vocal tract," Acta Acust. united Acust. **93**(6), 1046–1056.

Bond, Z. S., and Moore, T. J. (**1990**). "A note on loud and Lombard speech," in *Proceedings of the First International Conference on Spoken Language Processing (ICSLP)*, November 18–22, Kobe, Japan.

Bourne, T., and Garnier, M. (**2012**). "Physiological and acoustic characteristics of the female music theater voice," J. Acoust. Soc. Am. **131**, 1586–1594.

Bourne, T., Garnier, M., and Samson, A. (**2016**). "Physiological and acoustic characteristics of the male music theatre voice," J. Acoust. Soc. Am. **140**, 610–621.

Epps, J., Smith, J. R., and Wolfe, J. (**1997**). "A novel instrument to measure acoustic resonances of the vocal tract during speech," Meas. Sci. Technol. **8**, 1112–1121.

Garnier, M., Henrich, N., Smith, J., and Wolfe, J. (**2010**). "Vocal tract adjustments in the high soprano range," J. Acoust. Soc. Am. **127**, 3771–3780.

Garnier, M., Ménard, L., and Alexandre, B. (**2018**). "Hyper-articulation in Lombard speech: An active communicative strategy to enhance visible speech cues?" J. Acoust. Soc. Am. **144**, 1059–1074.

Geumann, A. (**2001**). "Vocal intensity: Acoustic and articulatory correlates," in *Proceedings of the 4th Conference on Motor Control*, June 13–16, Nijmegen, Netherlands.

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., and Perkins, W. H. (**1988**). "Relationship between changes in voice pitch and loudness," J. Voice **2**, 118–126.

Hanna, N., Smith, J., and Wolfe, J. (**2016**). "Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation," J. Acoust. Soc. Am. **139**, 2924–2936.

Henrich, N., d'Alessandro, C., Doval, B., and Castellengo, M. (**2004**). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," J. Acoust. Soc. Am. **115**, 1321–1332.

Henrich, N., Kiek, M., Smith, J., and Wolfe, J. (**2007**). "Resonance strategies used in Bulgarian women's singing style: A pilot study," Logoped. Phoniatr. Vocol. **32**, 171–177.

Henrich, N., Smith, J., and Wolfe, J. (**2011**). "Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones," J. Acoust. Soc. Am. **129**, 1024–1035.

Hoemeke, K. A., and Diehl, R. L. (**1994**). "Perception of vowel height: The role of F_1–F_0 distance," J. Acoust. Soc. Am. **96**, 661–674.

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," J. Acoust. Soc. Am. **84**, 511–529.

Isshiki, N. (**1964**). "Regulatory mechanism of voice intensity variation," J. Speech Hear. Res. **7**, 17–29.

Joliveau, E., Smith, J., and Wolfe, J. (**2004**). "Vocal tract resonances in singing: The soprano voice," J. Acoust. Soc. Am. **116**, 2434–2439.

Lienard, J. S., and Di Benedetto, M. G. (**1999**). "Effect of vocal effort on spectral properties of vowels," J. Acoust. Soc. Am. **106**, 411–422.

Morrison, M. D., and Rammage, L. A. (**1993**). "Muscle misuse voice disorders: Description and classification," Acta Otolaryngol. **113**, 428–434.

Pabst, F., and Sundberg, J. (**1993**). "Tracking multi-channel electroglottograph measurement of larynx height in singers," Scand. J. Logopedics Phoniatrics **18**(4), 143–152.

Plant, R. L., and Younger, R. M. (**2000**). "The interrelationship of subglottic air pressure, fundamental frequency, and vocal intensity during speech," J. Voice **14**, 170–177.

Raitio, T., Suni, A., Pohjalainen, J., Airaksinen, M., Vainio, M., and Alku, P. (**2013**). "Analysis and synthesis of shouted speech," in *Interspeech 2013*, August 25–29, Lyon, France, pp. 1544–1548.

Raphael, B. N., and Scherer, R. (**1987**). "Voice modifications of stage actors: Acoustic analyses," J. Voice **1**, 83–87.

R Core Team (**2013**). "R: A language and environment for statistical computing" (R Foundation for Statistical Computing, Vienna, Austria).

Rostolland, D. (**1982a**). "Acoustic features of shouted voice," Acta Acust. **50**, 118–125.

Rostolland, D. (**1982b**). "Phonetic structure of shouted voice," Acta Acust. **51**, 80–89.

Roy, N., Nissen, S. L., Dromey, C., and Sapir, S. (**2009**). "Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy," J. Commun. Disord. **42**, 124–135.

Schulman, R. (**1989**). "Articulatory dynamics of loud and normal speech," J. Acoust. Soc. Am. **85**, 295–312.

Smith, J. R. (**1995**). "Phasing of harmonic components to optimize measured signal-to-noise ratios of transfer functions," Meas. Sci. Technol. **6**, 1343–1348.

Sundberg, J. (**1975**). "Formant technique in a professional female singer," Acta Acust. united Acust. **32**(2), 89–96.

Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois University, DeKalb, IL).

Sundberg, J., Lã, F., and Gill, B. (**2011**). "Professional male singers' formant tuning strategies for the vowel /a/," Logoped. Phoniatr. Vocol. **36**, 156–167.

Sundberg, J., and Nordenberg, M. (**2006**). "Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech," J. Acoust. Soc. Am. **120**, 453–457.

Swerdlin, Y., Smith, J., and Wolfe, J. (**2010**). "The effect of whisper and creak vocal mechanisms on vocal tract resonances," J. Acoust. Soc. Am. **127**, 2590–2598.

Titze, I. R. (**1989**). "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. **85**, 901–906.

Titze, I. R. (**2001**). "Acoustic interpretation of resonant voice," J. Voice **15**, 519–528.

Titze, I. R. (**2004**). "A theoretical study of f_0-f_1 interaction with application to resonant speaking and singing voice," J. Voice **18**, 292–298.

Titze, I. R. (**2008**). "Nonlinear source–filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–2749.

Titze, I. R., and Sundberg, J. (**1992**). "Vocal intensity in speakers and singers," J. Acoust. Soc. Am. **91**, 2936–2946.

Traunmüller, H. (**1981**). "Perceptual dimension of openness in vowels," J. Acoust. Soc. Am. **69**, 1465–1475.

Traunmüller, H., and Eriksson, A. (**2000**). "Acoustic effects of variation in vocal effort by men, women, and children," J. Acoust. Soc. Am. **107**, 3438–3451.

Vurma, A. (**2022**). "Amplitude effects of vocal tract resonance adjustments when singing louder," J. Voice **36**(2), 292.e11–292.e22.

Xue, Y., Marxen, M., Akagi, M., and Birkholz, P. (**2021**). "Acoustic and articulatory analysis and synthesis of shouted vowels," Comput. Speech Lang. **66**, 101156.