JSLHR

Research Article

1951

Vowel Formants in Normal and Loud Speech

Laura L. Koenig[a,b,c] and Susanne Fuchs[c]

Purpose: This study evaluated how 1st and 2nd vowel formant frequencies (F1, F2) differ between normal and loud speech in multiple speaking tasks to assess claims that loudness leads to exaggerated vowel articulation.
Method: Eleven healthy German-speaking women produced normal and loud speech in 3 tasks that varied in the degree of spontaneity: reading sentences that contained isolated /i: a: u:/, responding to questions that included target words with controlled consonantal contexts but varying vowel qualities, and a recipe recall task. Loudness variation was elicited naturalistically by changing interlocutor distance. First and 2nd formant frequencies and average sound pressure level were obtained from the stressed vowels in the target words, and vowel space area was calculated from /i: a: u:/.
Results: Comparisons across many vowels indicated that high, tense vowels showed limited formant variation as a function of loudness. Analysis of /i: a: u:/ across speech tasks revealed vowel space reduction in the recipe retell task compared to the other 2. Loudness changes for F1 were consistent in direction but variable in extent, with few significant results for high tense vowels. Results for F2 were quite varied and frequently not significant. Speakers differed in how loudness and task affected formant values. Finally, correlations between sound pressure level and F1 were generally positive but varied in magnitude across vowels, with the high tense vowels showing very flat slopes.
Discussion: These data indicate that naturalistically elicited loud speech in typical speakers does not always lead to changes in vowel formant frequencies and call into question the notion that increasing loudness is necessarily an automatic method of expanding the vowel space.
Supplemental Material: https://doi.org/10.23641/asha.8061740

Loudness modulation is a natural part of typical speech production, and targeting loud speech is also used in clinical intervention for persons with motor speech disorders. Elucidating the ways in which speech changes from normal to loud conditions is therefore important for basic research and may provide insight into the physiological mechanisms underlying therapeutic methods. As described below, reports of changes in vowel formant frequencies and supraglottal kinematics[1] in loud speech have led to claims that loudness modifies speech behavior across all levels of the production system (respiratory, phonatory, articulatory). A review of this work shows, however, that results have varied considerably across studies, as have methodological factors such as speech materials and speaking tasks. This article evaluates the degree to which vowel formants change in louder speech produced by healthy women to assess the generality of loudness-related effects across vowel types, speaking tasks, and speakers. We first provide a brief overview of the nature of speech production changes in loud speech and theoretical perspectives on the effects of loudness. This is followed by a more detailed review of the empirical support for the idea that loud speech yields widespread changes in speech production, particularly as measured by vowel formant frequencies, both in typical speakers and those with motor speech disorders.

General Characteristics of Loud Speech and Methods of Study

Traditionally, increased loudness has primarily been associated with greater respiratory system displacements and higher subglottal pressures (e.g., Ladefoged & McKinney, 1963). Louder speech may also differ in laryngeal setting (e.g., Finnegan, Luschei, & Hoffman, 2000) and usually

[a]Adelphi University, Garden City, NY
[b]Haskins Laboratories, New Haven, CT
[c]Leibniz-Centre General Linguistics (ZAS), Berlin, Germany
Correspondence to Laura L. Koenig: koenig@haskins.yale.edu

Editor-in-Chief: Julie Liss
Editor: Bharath Chandrasekaran
Received February 2, 2018
Revision received July 1, 2018
Accepted December 3, 2018
https://doi.org/10.1044/2018_JSLHR-S-18-0043

[1]We remind readers that articulation and acoustics are interrelated (e.g., Fant, 1971), and authors frequently use formants as a way of gaining insight into articulation.

Disclosure: The authors have declared that no competing interests existed at the time of publication.

has a higher fundamental frequency than typical speech (e.g., Lieberman, Knudson, & Mead, 1969). Over the last few decades, a growing number of studies have also reported differences in supraglottal articulation and vowel formant frequencies across loudness conditions. To facilitate subsequent discussion and comparison across studies, Table 1 summarizes methods and results of loud speech studies. The table shows that authors have used a variety of methods to elicit loud speech: varying interlocutor distance, explicit instructions (including the use of loudness in therapy), and employing noise to induce greater loudness (the Lombard effect; see Lane & Tranel, 1971, and citations therein). Speech materials have also varied considerably.

Theoretical Perspectives on Loud Speech

Early studies of relationships between speech articulation and acoustics indicated that greater oral apertures (which correlate with higher values of the first formant, F1) led to greater acoustic intensity (Fant, 1971; Lindblom & Sundberg, 1971). That is, more open articulatory postures may, along with changes in laryngeal setting and respiratory drive, provide an additional mechanism for increasing loudness (cf. Dromey & Ramig, 1998). This is presumably what yields the "intrinsic intensity" effect (Lehiste, 1970), whereby lower vowels have greater intensities on average.

More recent authors have attempted to explain how/why loudness might modify speaker behavior in a system-wide fashion (i.e., affect respiratory, laryngeal, and supralaryngeal behavior simultaneously). Schulman (1989, p. 310) proposed that loud speech was analogous to a "natural bite block" condition, that is, a form of perturbation that may provide insight into overall articulatory organization. Dromey, Ramig, and Johnson (1995, pp. 761–762) characterized loud speech as "a naturally occurring scaling transformation, which modifies the activity of all muscles in the articulatory linkage," analogous to the extensive alterations observed with changes in speech rate and stress. Similarly, Dromey and Ramig (1998, p. 1014) referred to sound pressure level (SPL) as a "global variable…that has an impact across the speech subsystems." Fox, Morrison, Ramig, and Sapir (2002) hypothesized that loudness could lead not only to increased amplitudes of articulatory movement but also to improved coordination in the orofacial system. Sapir et al. (2003) further speculated that loudness effects could be related to limbic system or reflexive actions.

All of these characterizations presuppose that loudness does, in fact, affect supraglottal articulation (along with the expected respiratory and laryngeal adjustments). A few authors have provided counterevidence to the claim that loud speech necessarily has widespread effects on speech production (Tasko & McClean, 2004; Tjaden & Wilding, 2005), but to date, such reports have not been numerous. The primary purpose of this work was to test the hypothesis that louder speech leads reliably to changes in vowel formant frequencies in healthy speakers, considering different vowel contexts and speaker tasks. Formant frequency differences between loud and normal speech would provide evidence for differences in supralaryngeal articulation.

Loud Speech in Typical Speakers

Articulation

In a widely cited early study, Schulman (1989) assessed jaw and lip movements in shouted speech, sampling across the Swedish vowel system. The data showed more extreme jaw lowering and lip separation as loudness increased. The changes were greatest for low vowels, such that oral aperture differences across vowels were maintained in loud speech (i.e., low vowels still had greater apertures than high ones). Increased amplitudes were also associated with higher articulatory velocities. Subsequent studies of single low vowels have similarly demonstrated larger displacements and velocities of lip movements in louder speech (Darling & Huber, 2011; Dromey & Ramig, 1998; Huber & Chandrasekaran, 2006), and Mefferd and Green (2010) found larger tongue dorsum displacements of an /iə/ sequence in loud speech compared to normal. Finally, Tasko and McClean (2004) reported that louder speech had increased movement speed and displacement compared to typical. Thus, the literature consistently shows that loud speech affects aspects of supralaryngeal kinematics, although speakers may differ in the extent of these changes (see Dromey & Ramig, 1998; Schulman, 1989).

Acoustics

Lower positions of the jaw and greater oral apertures in loud speech would be expected to correlate with higher values of F1 (Lindblom & Sundberg, 1971), and such findings have been reported. For example, three studies of low vowels (Huber & Chandrasekaran, 2006; Huber, Stathopoulos, Curione, Ash, & Johnson, 1999; Traunmüller & Eriksson, 2000) obtained higher F1s in loud than typical speech. The magnitude varied greatly across studies, however (cf. Table 1). Also, despite observing an overall F1 increase for men, women, and children, Huber et al. (1999) found no F1 difference in loud speech produced by women. Such variation could arise from multiple sources, including the nature of the speaking task and the degree to which speakers actually changed their loudness levels.

Not many studies of loud speech have explored whether F1 changes in loud speech differ across vowels, and the existing data show mixed results (see Table 1). Bond, Moore, and Gable (1989) found that F1 changes varied across vowels in the order /i ɑ/ < /æ/ < /u/, but Tartter, Gomes, and Litwin (1993) obtained no main effect of noise on F1 frequencies in either of their two speakers and no clear patterns across vowels. Liénard and Di Benedetto's (1999) data did not show a significant Distance × Vowel interaction, implying that all vowels were equally affected by loudness changes.

Results for the second formant (F2) are quite varied (see Table 1). For example, whereas Huber and Chandrasekaran (2006) observed higher F2s for /ɑ/ in one of their loudness conditions, Huber et al. (1999) found no significant effects of loudness on /ɑ/ F2s. Bond et al.'s (1989) data showed slight

Table 1. Summary of studies reporting changes in supraglottal kinematics and formants in loud speech, following the order of their coverage in the literature review.

Citation	Speakers	Instrumentation	Speech conditions	Vowels assessed	Kinematic changes	F1 changes	F2 changes	Individual data?[a]
Schulman (1989)	3 men, 1 woman	Magnetometry; acoustics	Shouted speech targeted at 90 dB[b]	12 vowels in ib_b context	Lower jaw positions, greater lip aperture	n/a[c]	n/a	Yes
Darling & Huber (2011)	9 elderly adults[d]	Optotrak; acoustics	Noise of 70 dBA; twice and 10 dB[b] louder than comfortable	/ɑ/ in *Bobby*	Higher opening displacements, velocities for lower lip and jaw	n/a	n/a	No
Dromey & Ramig (1998)	5 men, 5 women	Strain gauge[e]; acoustics	Half, twice normal loudness[b]	/æ/ in *sapapple*	Higher displacements and velocities of lip opening and closing	n/a	n/a	Yes
Huber & Chandrasekaran (2006)	15 men, 15 women	Optotrak; acoustics	Noise of 70 dBA; twice and 10 dB[b] louder than comfortable	/ɑ/ in *Bobby*	Higher opening and closing displacements, velocities for lower lip and jaw[f]	ca. 80 Hz higher[g]	ca. 80 Hz higher at +10 dB	No
Mefferd & Green (2010)	5 men, 5 women	Magnetometry; acoustics	Twice normal loudness	/iə/ in *Mia*	Larger lingual displacements	n/a[h]	n/a[h]	No
Tasko & McClean (2004)	15 men	Magnetometry; acoustics	Half, twice comfortable loudness[b]	/æ/ in *bad*; vowels in connected speech	Higher displacements and velocity	n/a	n/a	No
Huber et al. (1999)	10 men, 10 women[i]	Acoustics	+10 dB[b] over comfortable	sustained /ɑ/	n/a	35 Hz higher but see text	Noise NS[j]	No
Traunmüller & Eriksson (2000)	6 men, 6 women[i]	Acoustics	Vary distance	/a/ in 3 words	n/a	ca. 200 Hz higher for men, 400 Hz for women[k]	n/a[l]	No
Bond et al. (1989)	4 men	Acoustics	Noise	/i ɑ æ u/ from 10 disyllabic words	n/a	25–125 Hz higher, with /i ɑ/ < /æ/ < /u/[m]	Slight fronting for /u/, retraction for /i/[m]	No
Tartter et al. (1993)	2 women	Acoustics	Noise: 35, 60, 80 dB	Vowels in digits "zero"–"nine"	n/a	Noise NS	Speaker 1: Noise significant for "0," "1," "3," "8," "9" mostly at 80 dB. Speaker 2: Differences for "2" and "9"	Yes
Liénard & Di Benedetto (1999)	5 men, 5 women	Acoustics	Vary distance	9 (oral) vowels of French	n/a	ca. 30–100 Hz higher across vowels	ca. 50 Hz max; variable across vowels	No
Junqua (1993)	5 men, 5 women	Acoustics	Noise	Vowels in digits "zero"–"nine"; alphabetic letters; 13 words	n/a	42–113 Hz higher for vowels, sonorants	Higher for women; lower for men	No
Lu & Cooke (2008)	4 men, 4 women	Acoustics	Noise and babble	/i ɪ e u/ from 4 words in sentences	n/a	"up to 100 Hz" higher (p. 3266)	Increased for /i/ and /ɪ/	No

(*table continues*)

Table 1. *(Continued).*

Citation	Speakers	Instrumentation	Speech conditions	Vowels assessed	Kinematic changes	F1 changes	F2 changes	Individual data?[a]
Dromey et al. (1995)	1 man with PD[n]	Acoustics[e]	Pre–post LSVT	/ɪ ɛ æ u o͡ʊ e͡ɪ a͡ɪ/ in 12 words	n/a	Not reported	F2 transitions ca. 100 Hz larger; most changes in high front vowels or offglides	n/a
Sapir et al. (2003)	1 woman with cerebellar ataxia	Acoustics	Pre–post LSVT	All vowels in 3 sentences; point vowels in *key*, *stew*, *Bobby*	n/a	Higher: ca. 50–75 Hz for high vowels, 200 Hz for /ɑ/	Higher for sentences; lower /u/, higher /i/ values in target words	n/a
Sapir et al. (2007)	14 typical; 29 with PD, split between LSVT and untreated	Acoustics	Pre–post LSVT	Point vowels in *key*, *stew*, *Bobby*	n/a	NS	Lower for /u/ in treated group	No
Wenke et al. (2010)	26 with dysarthria, split into LSVT and traditional therapy groups	Acoustics	LSVT, traditional therapy	/i æ ɑ u/ in b_t context	n/a	NS	NS	No
Tjaden & Wilding (2004)	15 typical; 12 with PD; 15 with MS[o]	Acoustics	Habitual and 2× normal loudness	/i ɑ æ u a͡ɪ e͡ɪ/ from reading passage	n/a	Averages ca. 25 Hz higher overall; max. ca. 80 Hz in speakers with MS[p]	Averages ca. 20 Hz higher for /i æ/, lower for /ɑ u/[o] Max. effects ca. 75 Hz Larger extents for /e͡ɪ/; ca. 50% of speakers showed the effect	No[q]

Note. The table does not include work on linguistic factors (e.g., stress and accent) that may involve increased loudness as one effect. It also excludes clear speech studies, but see text for brief coverage of results from that body of literature. Study designs include (a) explicitly asking speakers to target louder speech; (b) using distance to induce louder speech; (c) Lombard speech, that is, speaking under noise; and (d) using LSVT methods. Some studies also included rate changes; in the name of brevity, those results are not reported here. The columns on kinematic and formant changes express differences for louder speech relative to typical. LSVT = Lee Silverman Voice Treatment.

[a]Considering only the measures of interest at present, namely articulatory displacements and formant frequencies. See text for details of individual speakers. [b]VU/SPL meter used for loudness feedback. [c]Schulman (1989) speaks of recording acoustics but presents no formant data. The abstract of Schulman (1985) states that loud speech showed increased values of F1, but magnitudes are not available, and it is not clear whether the acoustic data are drawn from the entire data set of Schulman (1989). n/a = not applicable. [d]Study also included nine speakers with PD. Their kinematic changes were similar to those of typical speakers but were not as straightforwardly related to SPL changes. [e]Only reporting methods and results for supraglottal articulators. [f]One result (lower lip opening velocity at +10 dB) did not reach significance. [g]Averaging across conditions. [h]The authors reported a larger Euclidean distance between /i/ and /ə/ in F1–F2 space but separate F1 and F2 differences are not recoverable. [i]Study also included children. [j]NS = not significant. [k]Not reporting data on whispered conditions, that is, utterances produced at very close range. [l]The authors mention that they had observed variation across vowels in F2 changes in past work but provide no details. [m]Qualitative; the authors do not report statistical findings. [n]PD = Parkinson's disease. [o]MS = multiple sclerosis. [p]Statistical analysis was based on vowel space areas. F1 and F2 changes are estimated from the figures. [q]The authors do report on the number of speakers in each group who conformed to the patterns shown in the group-level statistics.

centralization for high vowels, whereas Lu and Cooke (2008) reported significant increases in F2 for high front vowels. The speakers in Tartter et al. (1993) had scattered F2 effects that did not yield straightforward interpretation, and Junqua's (1993) data showed no significant vowel differences at all. In short, in typical speakers, the literature does not suggest consistent differences for F2 as a function of loudness.

Formant Changes in Loud Speech in Dysarthria

Therapy methods for speakers with dysarthria often target changes in rate or loudness (e.g., Dromey & Ramig, 1998; Tjaden, Lam, & Wilding, 2013). If speaking more loudly yields changes in multiple speech systems (respiratory, laryngeal, supralaryngeal), it could function as an effective means of achieving wide-scale changes in speech production. One intervention method that has received considerable attention is the Lee Silverman Voice Treatment (LSVT; Ramig, Bonitati, Lemke, & Horii, 1994). LSVT is an intensive protocol that primarily encourages speakers to produce louder, more effortful speech (Ramig, Countryman, Thompson, & Horii, 1995). It has most often been used in intervention for Parkinson's disease (PD), where patients frequently show reduced vocal loudness and phonatory abnormalities (e.g., Fox et al., 2002). Studies of LSVT have considered many aspects of speech production; in line with the focus of this article, the following review is restricted to studies that included measures of vowel formants. Whereas investigations of loud speech in typical speakers have evaluated changes in F1 and F2 separately, much of this clinically oriented work has also assessed overall vowel space areas (VSA), in light of evidence that such measures may correlate with intelligibility (e.g., Bradlow, Torretta, & Pisoni, 1996; but see Tjaden & Wilding, 2004, for some caveats).

Two early case studies suggested that LSVT might have effects on vowel formants. Dromey et al. (1995) reported a posttherapy increase in averaged F2 transition extents and durations, particularly for high front vowels. Sapir et al. (2003) obtained higher F1 and F2 frequencies averaged over vowels and an expansion of F2 (more extreme values for /i/ and /u/ in loud speech compared to typical).

In a more extensive study, Sapir, Spielman, Ramig, Story, and Fox (2007) assessed formant values for speakers with typical neurological function and persons with PD. One PD group received LSVT, and the other was an untreated control group. Prior to therapy, the two PD groups had significantly higher F2s in /u/ compared to the typical group. After therapy, the treated group showed a significant decrease in /u/ F2s. Finally, Wenke, Cornwell, and Theodoros (2010) compared traditional dysarthria therapy (TRAD) with LSVT. Vowel spaces increased posttherapy for LSVT (but not TRAD); at the same time, the LSVT versus TRAD vowel space comparisons were not significant before or after therapy. F1 and F2 values of individual vowels did not differ pre–post therapy in either group.

Tjaden and Wilding (2004) studied loudness (and rate) manipulations in single-session (nontreatment) paradigms, investigating typical speakers and those with dysarthria. Increased VSAs in loud speech were only observed in the typical speakers. Group formant plots show that, qualitatively (cf. Table 1), (a) most vowels had small F1 increases in loud speech, (b) F2 differences between front and back vowels became slightly larger, and (c) F2 transition extents increased for /eɪ/ in some speakers. Overall, F1 and F2 changes varied as a function of vowel, speaker gender, and speaker group.

In summary, there is some evidence that loudness variation may affect formant values in speakers with dysarthria, but the results are not totally consistent. Some differences across studies may, again, reflect methodological variation, for example, differences in dysarthria severity and whether louder speech was elicited via an LSVT training program or under laboratory conditions in a single session.

Loud Speech and Clear Speech[2]

Although clear speech is usually treated as a distinct experimental manipulation, it may have, as one characteristic, increased loudness compared to typical speech (Picheney, Durlach, & Braida, 1986; Tjaden, Sussman, & Wilding, 2014). Of importance in the current context, some studies report that clear speech is associated with formant changes, which may include an overall F1 increase or an expanded F1 space, that is, higher high vowels and lower low vowels (Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988; Picheney et al., 1986). Picheney et al. (1986) also noted increased F2 dispersion (greater front–back differences) in clear speech than typical speech, but the lax vowels tended to show fronting rather than F2 expansion. Tjaden et al. (2013) found larger vowel spaces in clear, loud, and slow speech compared to habitual speech, with the largest changes in clear speech. Vowel space changes were more extreme for nonperipheral vowels than peripheral vowels. Thus, there is some overlap between the characteristics of loud and clear speech, but changes in F2 may be more common in clear speech than loud speech. Some results of this work also demonstrate differences between tense versus lax or peripheral versus nonperipheral vowels.

[2]Our impression is that the boundary between clear and loud speech is somewhat fuzzy, and deeper study is needed to elucidate how speech behaviors compare across these speaking styles. Indeed, speaker instructions in clear speech studies often mention noisy conditions or an interlocutor with hearing impairment (e.g., Picheney, Durlach, & Braida, 1985; Smiljanić & Bradlow, 2005; Tjaden et al., 2013), which could predispose speakers to increase loudness as one modification. Conversely, some studies of "loud speech" (e.g., Summers et al., 1988) instructed their speakers to increase clarity. It was beyond the purpose of this article to evaluate such differences in detail, and as stated in the text, we chose a method that did not explicitly prime speakers to change clarity. However, it is becoming evident that variations in speaker instructions can lead to measurable differences in speech production behavior (Huber, 2007; Lam & Tjaden, 2013; Lam et al., 2012; see also Gilbert, Victor, Chandrasekaran, & Smiljanic, 2013). Thus, researchers need to be specific about how they elicit changes in speaking styles, and readers should be mindful of possible effects of different instructions.

Methodological Considerations

As is evident in Table 1, most acoustic studies of loud speech have averaged data over multiple speakers. Yet, Tartter et al. (1993) observed numerous acoustic differences in how their two speakers changed from normal to loud conditions. Studies have differed in the degree of loudness variation they target, from modest (e.g., 5 dB differences) to shouting. Even for consistent instructions, the literature shows extensive variation in the degree to which speakers change loudness (Junqua, 1993; Liénard & Di Benedetto, 1999) and kinematic behavior (Dromey et al., 1995; Schulman, 1989) for a given task. For example, Junqua (1993) found a range of loudness change of 4–24 dB across speakers. Speaker variation is well documented—indeed, has received considerable attention—in clear speech studies (Ferguson & Kewley-Port, 2007; Hazan & Markham, 2004; Picheney et al., 1986; Summers et al., 1988). In contrast, research on loud speech has not widely explored speaker variation, at least for formants. As a result, it remains unclear to what extent there are "reliable and consistent" acoustic changes (Summers et al., 1988, p. 928) in loud speech and, therefore, whether louder speech uniformly serves as a natural method of "upscaling" all aspects of speech, including vowel production (e.g., Sapir et al., 2003, p. 388).

In addition, the existing literature has used a wide variety of speaking tasks (see Table 1). There is evidence that speaker behavior can vary as a function of task. Participants in Schmidt, Gelfer, and Andrews (1990) produced a larger range of intensity variation when asked to produce a loud-to-soft sequence than a soft-to-loud sequence. In Darling and Huber (2011), typical speakers and those with PD produced different levels of SPL change and kinematic behavior, depending on how loudness variation was elicited. Huber (2007) observed differences in respiratory behavior in reading as compared to monologue, and Tasko and McClean (2004) found greater articulatory durations, displacements, and velocities for a nonsense utterance compared to connected speech tasks. A better understanding of task effects would facilitate cross-study comparisons. Speech task effects may also reveal how attentional and cognitive demands affect speech production. Finally, one concern of therapeutic intervention is how well behavior generalizes across tasks.

Summary

The work summarized above leads us to the following conclusions.

1. Louder speech involves more open articulatory postures than typical speech, but the extent seems to be more extreme for low vowels (Schulman, 1989). This suggests possible differences in F1 changes across vowels.
2. Typical speakers tend to show higher F1 values in loud speech, but magnitudes vary across speakers and studies, and few authors have sampled widely across the vowel space. Most studies using multiple vowels had a very limited speech set (cf. Table 1), making it difficult to draw firm conclusions about how consistent effects may be across vowels. No apparent pattern emerges for F2 in typical speakers.
3. LSVT studies mostly find loudness-related differences in F2. The work of Tjaden and colleagues (Tjaden et al., 2013; Tjaden & Wilding, 2004) showed that vowel spaces may expand in loud (and clear) speech in some cases; across speakers and groups, there were some modest F1 and F2 changes that varied by vowel.
4. There are indications that speaker and task effects may be relevant, but these factors have not been widely explored.

Questions and Predictions

These observations lead us to the following questions.

1. How do formant frequencies differ between normal and loud speech when we sample widely across the vowel space, considering differences in vowel height and tense–lax status?
2. To what extent do changes associated with loud speech vary across speech tasks?
3. How much do speakers vary in their strategies for loud speech? Specifically, how consistent are loudness and task effects across speakers?
4. How well does intensity correlate with formant variations? In particular, are higher F1s in fact associated with higher SPLs?

We expect to see formant changes in loud speech for low vowels given the results on articulatory displacements and acoustics in such vowels summarized in Table 1. That is, we anticipate that, compared to nonlow vowels, low vowels in loud speech will have higher F1s, and that higher F1s, associated with more open articulatory positions, will correlate with increased loudness. Whether such a pattern generalizes to other vowel heights and across tense versus lax vowels is an open question. Based on past studies of loud (and clear) speech, we also expect that results may differ as a function of task and speaker.

Method

Speakers and Speaking Tasks

Data were obtained from 11 self-reported healthy female speakers of standard (northern) German. They ranged in age from 20 to 37 years (M = 26 years). Advertisements, posted around the Humboldt University community, invited healthy speakers for a study on speech physiology. As is typical for younger people in Germany, all participants had some familiarity with English as a second language and usually one other language. Participants completed three tasks (described below) in a fixed order. The Reading (1) and the Question–Answer (Q-A) (2) tasks were repeated so as to yield multiple repetitions of each target word. For the Pizza recipe task (3), speakers read different recipes, so that vocabulary overlap was only partial. The three different task types varied in the degree of spontaneity and the phonetic context: For read speech (1), we expected speakers to use a rather careful speaking style. This speaking task also elicited isolated vowels (albeit in a connected speech context), so that coarticulatory effects on the vowels should be minimal. The Q-A task (2) involved interaction between the experimenter and participant, and participants had to supply part of their response within a given sentence frame. All vowels were bounded by oral or nasal stop consonants (see below). The Pizza task required the most linguistic formulation from participants but also included a somewhat targeted vocabulary.

1. Reading sentences. The sentences, taken from Weirich and Simpson (2013), included acronyms containing the vowels [i: a: u:]:

   Sie fuhren letzte Woche zur IAA nach Frankfurt. ("They went last week to the IAA [i: a: a:] in Frankfurt").
   Wir wollen am Wochenende zur BII nach Hamburg. ("We want to go to the BII [be: i: i:] in Hamburg over the weekend")
   Sie fahren nächste Woche zur LUU nach Hannover. ("They're going next week to the LUU [ɛl u: u:] in Hannover.")

The results of this task should be most comparable to the sustained vowel conditions used in some past studies (cf. Table 1). By necessity, all target vowels in this task were tense, because lax vowels in German, like in English, require closed syllables (i.e., a coda consonant). That is, lax vowels in German (and English) are inherently subject to coarticulatory influences from both onset and coda consonants. This constraint also implies that sustained vowel tasks in English only include tense vowels.

2. Q-A task. Speakers provided answers to questions posed orally by the experimenter. For example:

   Question (Q): *Magst du X*? ("Do you like X?")
   Answer (A): *X mag ich, aber nicht Y* ("X I like, but not Y"; participants supplied their own Y).

The targeted words were all disyllabic, with a bilabial word onset and a medial alveolar consonant, and chosen to sample across the German vowel inventory: tense vowels [i: a: u: y:] and their lax counterparts [ɪ a ʊ ʏ]. The full word list is provided in Table 2.

3. Pizza recipe task. Participants read novel recipes for special pizzas (see Supplemental Material S4). After having a few minutes to read the recipe, participants put away the text and were asked to explain, using their own words, the steps for making the pizza to an assistant (the experimenter). The recipes had a certain amount of overlapping vocabulary (e.g., *Pizza*, *Mehl* ["flour"], *Salz* ["salt"], *Wasser* ["water"]). Participants were provided with two recipes for normal speech and two for loud speech.

Speakers first produced all three tasks in the normal voice condition and then in the loud voice condition in the following order: Reading (normal), Q-A (normal), Pizza (normal), Reading (loud), Q-A (loud), Pizza (loud). Because starting with loud speech could have had an impact on normal speech, the normal condition was always produced first to establish a baseline for each speaker.

Loudness variation was elicited naturalistically. In the first (normal) condition, the experimenter stood a few feet from the participant and spoke to her at a conversational loudness; in the second (loud) condition, the experimenter went into an adjacent room (ca. 5 m away), visible through a glass window, and spoke loudly to elicit the tasks. The Q-A and Pizza tasks both involved ongoing verbal interaction between the participant and the experimenter.

Recording

The full experimental protocol included simultaneous recordings of acoustics, intraoral pressure, electroglottography, and respiratory displacements. For this article, we report acoustic data only. These signals were obtained using a standing super cardio condenser microphone (Sennheiser HKH50 P48) coupled with a preamplifier. The microphone was positioned approximately 25 cm from the speaker's mouth. Speakers were seated during the experimental procedure to limit postural and positional variation.

Data Processing

The stressed vowels in the target words/vowels (e.g., /i:/, /a:/, *Mieten*, *Paten*, repeated words in the pizza recipes) were manually labeled in Praat (Version 6.0.27; Boersma & Weenink, 2016). Vowel onsets and offsets were identified on the basis of formant changes (for voiceless obstruents) or formant and intensity changes (for other consonant types). Words in the Pizza task were selected on a per-speaker basis, with the provision that the word was produced at least once in both normal and loud conditions. The final data set for all speakers was 3,110 tokens (674, 1,699, and 737 productions in the Reading, Q-A, and Pizza tasks, respectively; a speaker-by-speaker breakdown is provided in Supplemental Material S1). Formant and SPL measures were obtained automatically in Praat. Formant values were subsequently reviewed and hand-corrected with reference to the spectrogram, changing Praat parameters from their default settings (maximum value = 5500 Hz, window length = 25 ms, number of formants = 5) when spurious values were obtained.[3]

Final dependent measures were as follows:

(a) SPL, calculated as an average over the duration of each labeled vowel;
(b) F1 and F2, taken at the midpoint of the target vowel;

[3]A reviewer asked why we chose to report data in hertz rather than in Bark. The reason for this was because we were interested in the acoustics of speech production rather than in the perceptual ramifications of any changes. Classic studies of acoustic–articulatory relationships have typically reported data in hertz; see, for example, Stevens and House (1955), Fant (1971), and Lindblom and Sundberg (1971).

Table 2. Target words in the Question–Answer task.

Vowel	i:	ɪ	a:	a	u:	ʊ	y:	ʏ
Words	Mieten “rents” Pita “pita”	Mitte “middle” Pizza “pizza”	Mate “mate” (tea) Paten “godparents”	Paddeln “to canoe” Pasta “pasta”	Pudel “poodle” Pute “turkey”	Butter “butter” Pudding “pudding”	Büsten “busts” Büsum “Büsum” (an island)	Mützen “caps” München “Munich”

(c) a measure of VSA, proposed by Sapir, Ramig, Spielman, and Fox (2010), as a global indication of vowel space useful in assessing gains from therapy:

$$\mathrm{VSA} = \mathrm{ABS}((F1_i * (F2_a - F2_u) + F1_a * (F2_u - F2_i) + F1_u * (F2_i - F2_a))/2) \quad (1)$$

All statistical analyses were carried out in R (Version 3.4.1; R Core Team, 2017) using the *RStudio* editor (1.0.153), with the packages *ggplot2* for graphical illustration, *lme4* (Version 1.1-15; Bates, Mäechler, Bolker, & Walker, 2015) to run linear mixed-model analyses, *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017) to obtain degrees of freedom and *p* values via the Satterthwaite approximation, and *sjstats* to calculate analyses of variance with their corresponding effect sizes (η^2). Condition (loud vs. normal) and task (Reading vs. Q-A vs. Pizza) and their interaction were chosen as independent factors, unless otherwise specified. Word and speaker by condition by task served as the random structure (the latter was included for speaker-specific slopes). Several models were run differing in the complexity of the random structure. Model comparisons were carried out using the Akaike information criterion. The model with the smallest Akaike information criterion value was selected as the final one. Diagnostics were run to ensure that the residuals of the chosen models were normally distributed. The threshold for significance was .051. A Bonferroni correction was used to control for potential Type I errors. For this purpose, all *p* values were multiplied by the number of models that were run (three for intensity and six for formants); we will refer to these as corrected *p* values.

Results

Intensity Manipulation

As shown in Figure 1, participants increased SPL for the loud condition in all three tasks. The best *lmer* model was the one with condition, task, and their interaction as fixed factors and speaker by condition by task as well as word as the random structure. Reading and normal speech condition served as reference levels. The results showed main effects of both independent factors, but no interaction. The difference between loud and normal speech was 7.7 dB in the Reading task (intercept = 62.3 dB, β = 7.7 dB, df = 36.6, t = 8.82, corrected $p < .001$), 8.2 dB in the Q-A task (intercept = 67.3 Hz, β = 8.2 dB, df = 34.3, t = 9.05, corrected $p < .001$), and 6.4 dB in the Pizza task (intercept = 63.3 dB, β = 6.4 dB, df = 36.5, t = 7.4, corrected $p < .001$).

Speakers produced utterances in the Q-A task with 4.92 dB greater intensity on average than in the Reading task (intercept = 62.3 dB, β = 4.92 dB, df = 70.1, t = 4.76, corrected $p < .001$) and 3.99 dB greater than in the Pizza task (intercept = 67.26, β = 3.99, df = 48.8, t = 5.11, corrected $p < .001$). Intensity did not differ between the Reading and Pizza tasks. Figure 1 also shows that speaker loudness levels for normal and loud speech varied considerably. For example, s8's normal speaking level averaged 70.6 dB across the three tasks, whereas for s1, it was 58.4 dB. In the loud condition, s8 showed an average increase of 4.7 dB, whereas s1 increased by 11.3 dB for loud speech.

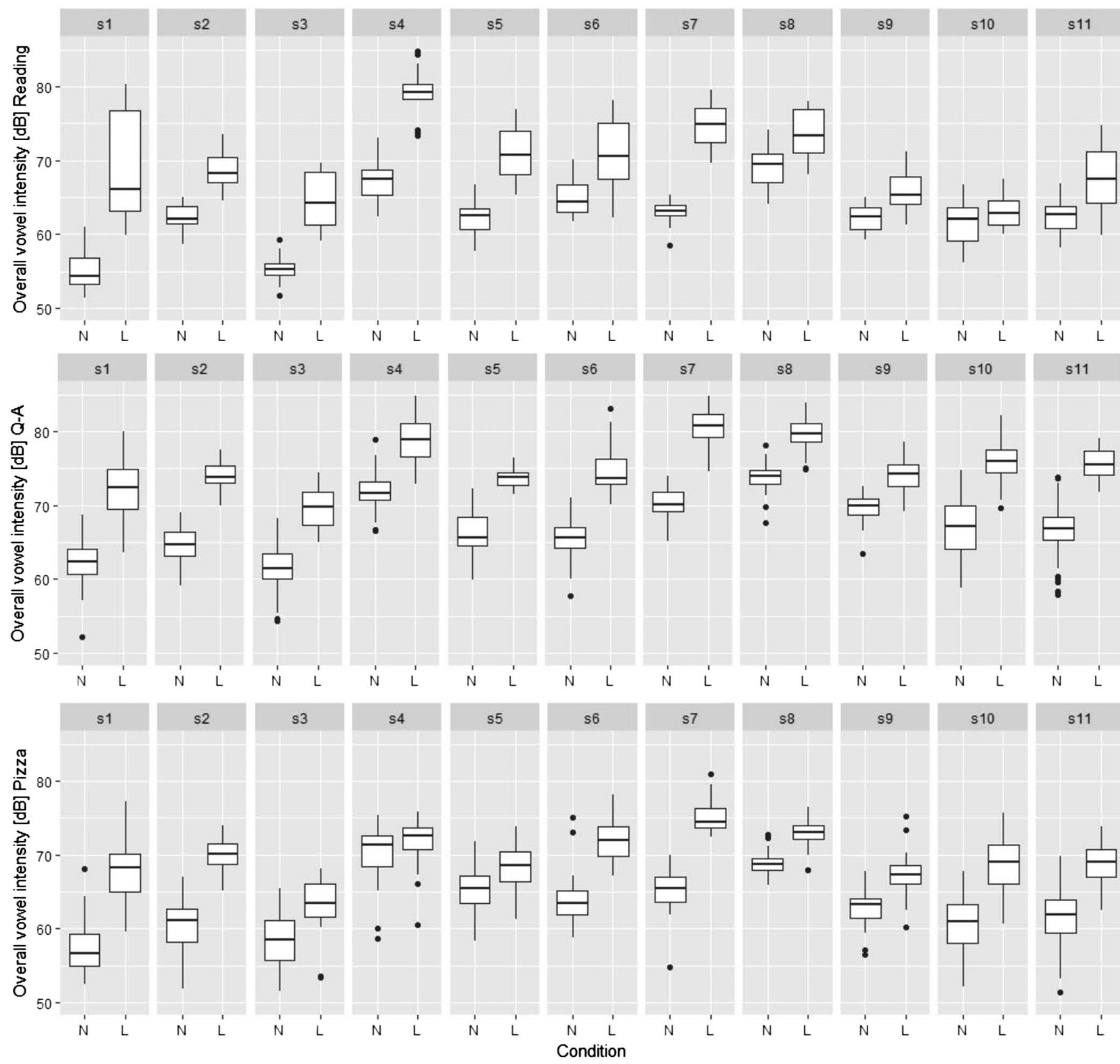
Vowel Differences in the Q-A Task

Our first question was how loudness affected different vowels. This section focuses on the Q-A task, which was designed to elicit a wide variety of vowel qualities, including multiple tense–lax pairs in a consistent consonantal context. Figure 2 shows the formant spaces for all speakers combined. Loudness effects were generally more extreme for low and lax vowels than high and tense vowels. For all vowels, loudness effects are clearest and most consistent for F1.

A linear mixed model with F1 as the dependent variable; condition (normal vs. loud), vowel height (high vs. low), and tenseness (tense vs. lax vowels) as independent variables; and speaker by condition, by vowel height, by tenseness as well as word as the random structure revealed significant main effects. The reference level was high lax vowels in normal condition. F1 varied significantly with vowel height (intercept = 435.8 Hz, β = 362.9 Hz, df = 65.7, t = 27.86, corrected $p < .001$, with higher values for low vowels) and tenseness (intercept = 435.8 Hz, β = –106.3 Hz, df = 58.4, t = –8.4, corrected $p < .001$, with lower values for tense vowels). Finally, there was a two-way interaction, with high lax vowels having a higher F1 than high tense vowels (intercept = 435.8 Hz, β = 218.7, df = 65.8, t = 11.8, corrected $p < .001$), whereas this was not the case for low vowels.

Condition also yielded significant effects. On average, F1 was 100 Hz higher (intercept = 435.8 Hz, β = 99.9 Hz, df = 40, t = 6.7, corrected $p < .001$) in loud than normal speech. The results also showed an interaction between condition and tenseness: The effect of loudness was smaller in tense than in lax vowels (intercept = 435.8 Hz, β = –78.4 Hz, df = 59.1, t = –4.36, corrected $p < .001$). For the high tense vowels, that is, /i:/, /u:/, /y:/, F1 did not differ significantly

Figure 1. Intensity data (dB) for all speakers (s1–s11) in the three tasks. From top to bottom: Reading, Question–Answer (Q-A), and Pizza. The horizontal bar in each box represents the median value. N = normal; L = loud.



between normal and loud conditions (intercept = 329.55 Hz, β = 21.52 Hz, *df* = 39.92, *t* = 1.45, corrected *p* = .936).

Running a similar model with F2 as the dependent variable did not yield a main effect of loudness or vowel height. There was an effect of tenseness (intercept = 1677 Hz, β = 146.6 Hz, *df* = 1693, *t* = 3.9, corrected *p* < .001), with higher F2 values for tense than lax vowels. An interaction was also found between tenseness and vowel height, with lower F2 values for tense low vowels than for lax low vowels (intercept = 1677.6 Hz, β = –310.8, *df* = 1694, *t* = –4.1, corrected *p* < .001).

Effects of Task, Vowel, and Speaker

The vowels [i: a: u:] were produced in all three tasks, so they provide the clearest indication of task-related differences for this data set. They also allow further exploration of vowel effects. Figure 3a shows VSA for the three tasks and two loudness conditions. Vowel space was calculated using the average formant values of the corner vowels for each speaker. Because the data set was fairly well balanced, an analysis of variance with condition and task as independent factors and VSA as the dependent variable

Figure 2. Formant variation across vowels in the Question–Answer (Q-A) task. Black = loud; gray = normal.

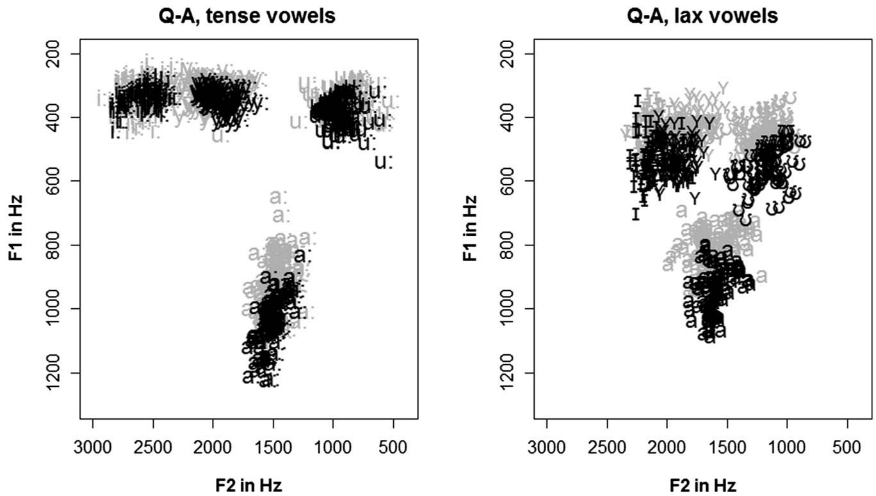
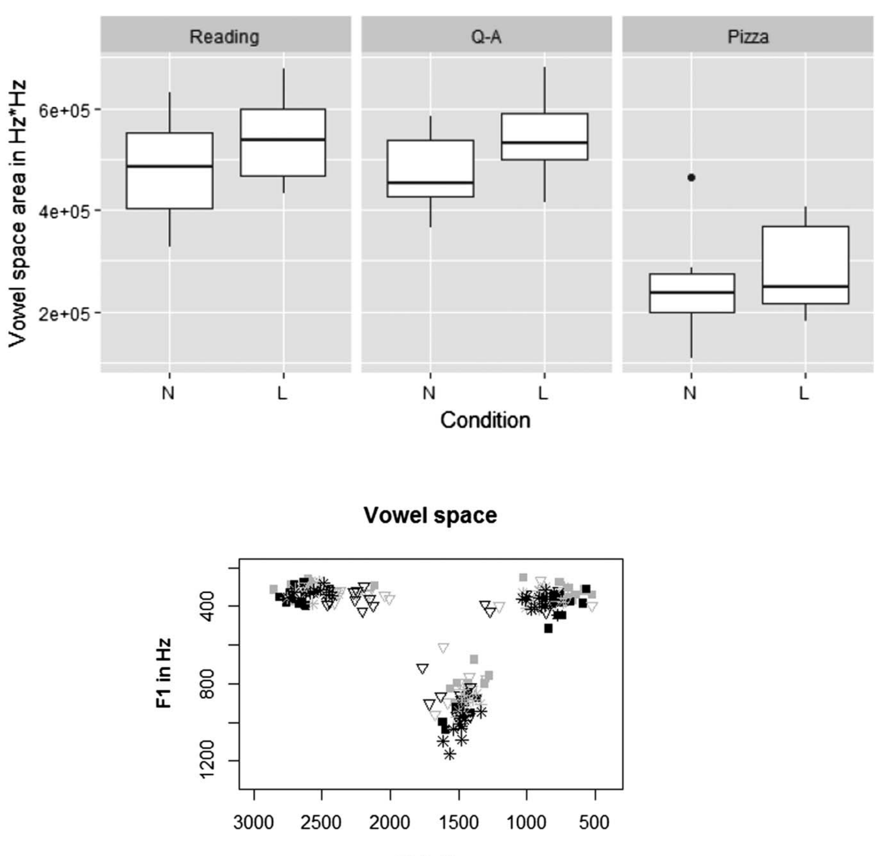


Figure 3. (a) Top: Vowel space areas (VSA; Hz^2). (b) Bottom: Vowel spaces, /i: a: u:/ only, all three tasks. Each symbol represents the mean for that vowel of a single speaker. All speakers are represented in the Reading and Question–Answer (Q-A) tasks; for the Pizza task, some speakers are not represented if they did not produce sufficient tokens for a given vowel. Black = loud; gray = normal. Reading: squares. Q-A: stars. Pizza: triangles.



was calculated. The results showed a significant effect of loudness, $F(1, 52) = 4.998$, $p = .0297$; vowel spaces were larger in loud speech. The effect of task was also significant, $F(2, 52) = 36.232$, $p < .001$. The Pizza task had the smallest overall vowel spaces. The Condition × Task interaction was not significant, $F(1, 52) = 0.231$, $p = .794$. Effect sizes were as follows: η^2 (condition) = .039, η^2 (task) = .558, η^2 (Task × Condition) = .0036. Thus, task effects were large compared to loudness effects.

Each speaker's averages for the three vowels are plotted in Figure 3b. The average F1 and F2 values in each task and loudness condition are provided in Supplemental Material S2. Qualitatively, F1 appears to increase with loudness in all cases, but the magnitude is largest for [a:]. Variations in F2 are small and vary with speaker, task, and vowel.

The nature of the vowel space reduction in the Pizza task is also evident in Figure 3b. On average, productions of /a:/ are raised and fronted somewhat. The largest F2 differences are seen for the high vowels: Many speakers show a centralization of /i:/, and a few show fronting of /u:/. These effects appear in both normal and loud speech. Formant differences between the Q-A and Reading tasks are small. However, the lowest productions of /a:/ appeared in the Q-A task, and the high vowels had more extreme F2 values in the Reading task than in other tasks.

To evaluate task differences formally, linear mixed models were run using all data (rather than averages) for the corner vowels. Data were split by vowel, because we were not interested in establishing the obvious differences in formant values among /a: i: u:/. Hence, for each vowel, a model was run with Condition × Task as fixed factors and Speaker × Task and Condition as well as Word as the random structure.

The /a:/ model revealed significant differences in F1 for condition and task, but no interaction. F1 was higher in the loud condition (Reading: intercept = 826.2 Hz, β = 106.2 Hz, $df = 31.87$, $t = 5.6$, corrected $p < 0.001$; Q-A: intercept = 911.12 Hz, β = 119.0 Hz, $df = 32.6$, $t = 6.24$, corrected $p < .001$; Pizza: intercept = 802.5 Hz, β = 69.15, $df = 75.3$, $t = 2.57$, corrected $p = .073$). Task effects were also observed for /a:/. F1s in the Reading task were significantly lower than those in the Q-A task (intercept = 911.1 Hz, β = –84.9 Hz, $df = 29.9$, $t = –4.99$, corrected $p < .001$), as were F1s in Pizza compared to Q-A (intercept = 911.1 Hz, β = –108.6 Hz, $df = 56.1$, $t = –4.91$, corrected $p < .001$). F1 did not differ between Reading and Pizza tasks.

For /i:/ and /u:/, the only significant effects were found for condition in the Reading task, with higher F1s in loud speech (for /i:/, intercept = 314.1 Hz, β = 30.96 Hz, $df = 33.1$, $t = 2.9$, corrected $p = .036$ and for /u:/, intercept = 334.2 Hz, β = 56.32 Hz, df =35.6, $t = 3.5$, corrected $p = .007$).

For F2, significant results were less frequent. For /a:/, no differences between normal and loud speech or different tasks were found. For /i:/, no differences were found between normal and loud speech, but F2 values were lower in the Pizza task than the Reading task (intercept = 2666 Hz, β = –317.9 Hz, $df = 3.61$, $t = –5.46$, corrected $p = .044$). Finally, /u:/ showed higher F2 values in the Q-A task compared to Reading (intercept = 643.5 Hz, β = 257.8 Hz, $df = 32.7$, $t = 5.9$, corrected $p < .001$) and likewise for the Pizza task compared to Reading (intercept = 643.5 Hz, β = 224.8 Hz, $df = 57.3$ $t = 3.9$, corrected $p = .001$).

To provide some insight into speaker variation, data from four participants are presented in Figure 4. These show that speakers differed in the magnitude of formant changes with loudness. In the Q-A task, for example, s3 and s11 produced rather large changes in F1 for /a:/, but changes for this vowel were more minor for s4 and s9. In the Reading task, s3 demonstrated centralized values for loud /i:/ and /u:/, which was not observed for the other three speakers.

Formant Variation and SPL

Increased oral apertures may represent a possible mechanism of increasing loudness. To what degree, then, does F1 correlate with SPL, across and within speakers? The data, given in Figure 5, indicate that the magnitude of the relationship varies widely, depending on the vowel quality.[4] In particular, the high tense vowels /i: u:/ in the Reading task and /i: y: u:/ in the Q-A task show little relationship between F1 and intensity, whereas the low and lax vowels—which, as seen above, showed more extreme F1 variation in loud speech—have a clearer relationship with loudness.

Figure 6 shows the data for each vowel in the Q-A task split by speaker. For the lax vowels, all slopes are positive. This is not the case for the tense vowels and especially the high ones /i: u: y:/. Thus, the general lack of a relationship between SPL and F1 for these vowels did not arise from combining data from multiple speakers.

Discussion

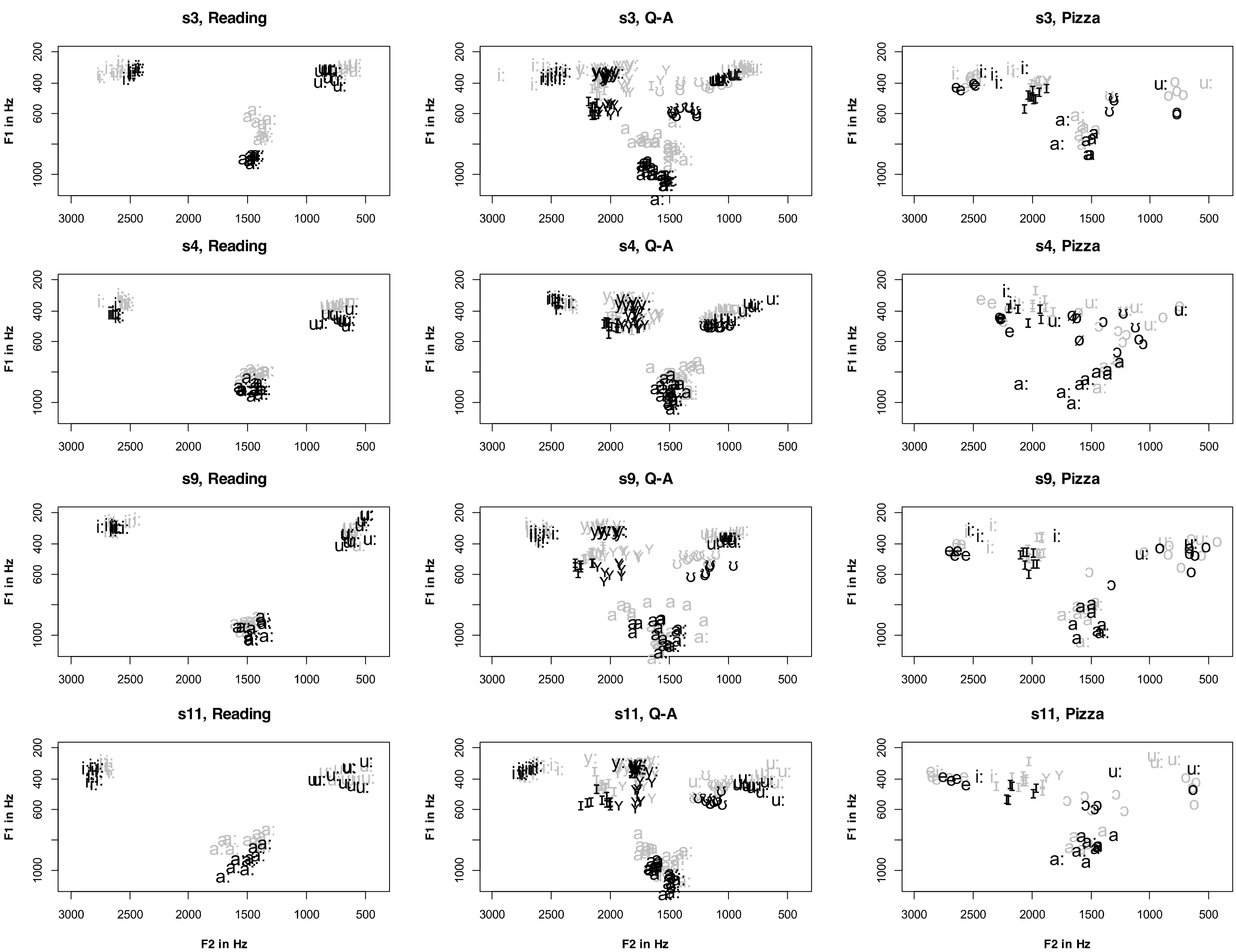
This work was designed to assess how loudness variation, elicited naturalistically in typical speakers, affected vowel formants. Our primary motivation was similar to that of previous studies (e.g., Dromey & Ramig, 1998; Tasko & McClean, 2004; Tjaden & Weismer, 1998) that sought to explore how speech behavior varies under manipulations that speakers encounter under everyday conditions. More specifically, we were interested in the degree to which increasing loudness yields an expansion of the vowel space, as predicted by the theoretical framework put forth by Dromey et al. (1995); see also, for example, Dromey and Ramig (1998) and Schulman (1989).

Vowel Differences

Schulman's (1989) study of shouted speech showed that changes in oral apertures were most extreme for

[4]Comparable plots for F2 are available in Supplemental Material S3. Most slopes were fairly flat, no clear pattern emerged across vowels or tasks, and indeed, the acoustic literature did not predict effects.

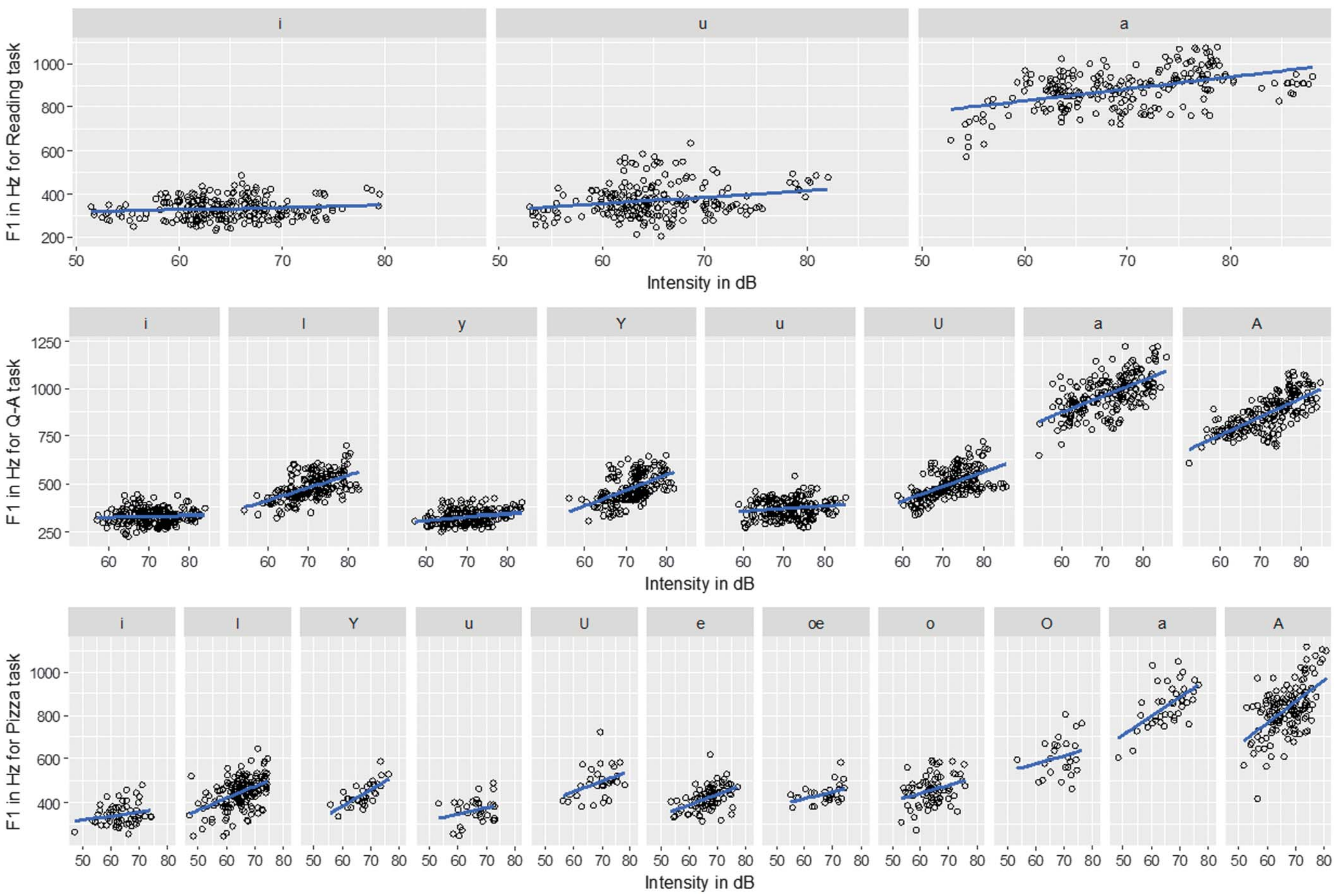
Figure 4. Vowel formant plots for four speakers (s3, s4, s9, s11), across tasks. Black = loud; gray = normal. Letters in upper case represent the lax vowels.



low vowels, which could suggest vowel differences in F1 patterns. The acoustic literature, on the other hand, was rather mixed as far as the consistency and magnitude of F1 changes across vowels in loud (if not necessarily shouted) speech and quite inconsistent as far as effects on F2.

In the current data, F1 did tend to increase in loud speech, but changes were minimal (in several cases, not significant) for high tense vowels. Loudness-related changes to F2 were both vowel and speaker specific. The F2 results suggest that group data, measuring across the vowel space, will typically not show a main effect of loudness for F2. Taking the results for F1 and F2 together, our data do not indicate that loud speech in typical speakers leads to a systematic, overall expansion of the vowel space. Insofar as F1 range increases, the effect in our data arises primarily from changes in low and lax vowels, in line with Schulman's (1989) articulatory data. The differences between the tense and lax high vowels are consistent with the results of Picheney et al. (1986) and are also reminiscent of Tjaden et al.'s (2013) observation of larger increases in VSA (albeit with considerable variability) in nonperipheral vowels than peripheral ones. Such cross-vowel differences may reflect a variety of factors: (a) production requirements, for example, coarticulation resistance arising from a raised tongue body (Recasens & Espinosa, 2009) and/or the use of tongue bracing in high tense vowels (cf. Strange et al., 2007); (b) perceptual characteristics, for example, /i/ has been reported to be one of the most well-recognized vowels (e.g., Syrdal & Gopal, 1986); (c) phonotactic patterns, such as syllable structure restrictions on lax vowels; and/or (d) linguistic effects, for example, the fact that lowering /i/ could make it confusable with other front vowels. Whatever the reason, it is clear in the current data that not all vowels behave the same in loud speech. Furthermore, the varying slopes for the intensity–F1 relationship across vowels

Figure 5. F1 and overall intensity, pooled across speakers, each task. Lax vowels are indicated by uppercase letters. Top: Reading. Middle: Question–Answer (Q-A). Bottom: Pizza task. Letters in upper case represent lax vowels.



suggest that increased articulatory opening and higher F1s may be an effective means of increasing loudness for some vowels, but not all.

Speaker Differences

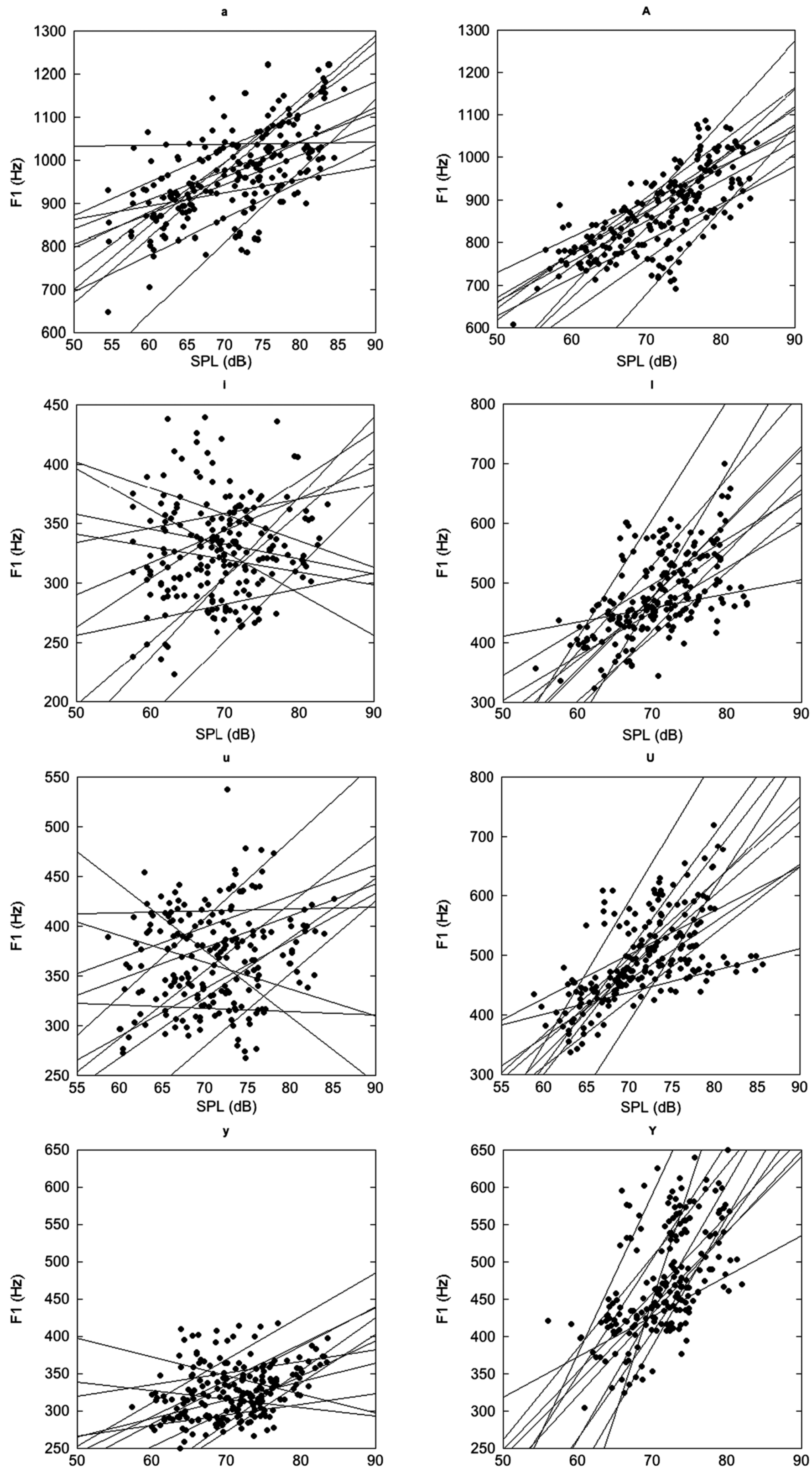
Although past work provided some evidence for speaker differences, the paucity of individual data made it difficult to assess the range of strategies that speakers may use for loud speech. The results here demonstrate considerable individual variation, a finding that echoes results on clear speech (Ferguson & Kewley-Port, 2007; Hazan & Markham, 2004). Speaker differences were also observed in Tjaden et al. (2013; see also Tjaden & Wilding, 2004): In their data, 33% of speakers did not show larger vowel spaces in loud speech (with similar percentages for participants with and without dysarthria). The cross-speaker variation observed in this study implies that there are multiple ways of speaking loudly. Authors studying speech rate (e.g., Byrd & Tan, 1996) have come to a similar conclusion: Speakers may adopt different strategies to produce speech at a faster (or slower) rate. Indeed, the possibility of multiple strategies for loud speech was recognized by Dromey and Ramig (1998). A general implication of this finding is that results based only on group-level analyses need to be treated with some caution, because they may not accurately capture the behavior of all individuals.

Task Differences and Comparisons Across Populations

The current results revealed several examples of task effects on vowel formants. This in itself is not surprising; numerous authors have observed that formants vary as a function of speech task (e.g., see the citations on clear speech above). A somewhat more interesting question is whether loudness effects on vowel formants differed across speech tasks (cf. Tasko & McClean, 2004). Whereas the global measure of VSA revealed no Condition × Task interaction,[5] more detailed analyses of the point vowels indicated loudness effects for high tense vowels only in the Reading task, that is, the one that presumably elicited the

[5]A recent paper by Whitfield, Dromey, and Palmer (2018) also points to limitations of simplified vowel space measures.

Figure 6. F1 by intensity for each speaker, Question–Answer (Q-A) task. Tense vowels on the left; lax vowels on the right.



most careful speech and the least coarticulatory variation —and which is the most similar to the sustained vowel tasks used in some past work. It could be that the Q-A task, in which vowels were always produced between stop consonants, had the effect of limiting changes in jaw position. Context effects do not appear to provide a straightforward explanation for the lack of effects in the Pizza task, where the analyzed vowels had a range of phonetic environments, including open syllables.

Comparisons within speaker (see Figure 4) provided a further indication that the behavior of particular vowels in louder speech was not always the same across tasks (e.g., see the patterns for /u:/ in s9). The slopes between intensity and F1 were also not always consistent across task; for example, whereas the slope of this relationship was rather flat for /i:/ in the Reading and Q-A task, it was positive for the Pizza task. In short, our data indicate that results obtained for one kind of speaking task may not be obtained in other types of speech material. This result supports the use of multiple speaking tasks to assess the consistency of articulatory and/or acoustic changes in loud speech.

Past authors have proposed that results from typical speakers may lend insight into therapeutic methods. For example, in a study of respiration, phonation, and supraglottal kinematics in typical speakers, Dromey and Ramig (1998) observed more consistent speech production changes and less interspeaker variability for loudness conditions compared to rate conditions. Based on these findings, those authors concluded that targeting loudness in therapy would yield more predictable results than targeting speech rate. Subsequent studies of LSVT (e.g., Baumgartner, Sapir, & Ramig, 2001; Sapir et al., 2003) have pointed to this article as support for the claim that loudness modulation affects supraglottal aspects of speech, along with the respiratory and laryngeal changes expected to arise from louder speech.

Despite these precedents in the literature, it remains a somewhat open question to what extent results from typical speakers can be generalized to those with dysarthria. Some studies suggest that such comparisons may not be straightforward. Darling and Huber (2011) observed that changes in articulatory displacements and velocities were more clearly related to changes in SPL in typical speakers than in those with PD. In a study of increased vocal intensity in older adults with and without PD, Matheron, Stathopoulos, Huber, and Sussman (2017) found voice quality changes in both groups, but the healthy adults "appeared to have better control of the laryngeal mechanism to make changes to their vocal intensity" (p. 507). Thus, neurotypical speakers may have access to at least as many options for changing speech behavior in loud conditions, if not more. It could be that a comparable increase in loudness for persons with motor speech disorders such as PD effectively requires a greater increase in speech production effort than in typical speakers, such that the appropriate comparison is not between varying loudness levels, but rather differences in effort. Indeed, some authors (e.g., Baumgartner et al., 2001; Ramig et al., 2001) refer to LSVT as targeting "effortful" speech, which may, in turn, lead to hyperarticulation in a manner similar to clear speech.[6]

As laid out in the introduction, the literature is rather consistent regarding how loudness affects phonation, respiration, and articulatory kinematics (i.e., the areas assessed by Dromey & Ramig, 1998). Furthermore, multiple studies (e.g., Ramig et al., 1995; Sapir et al., 2007) have demonstrated that LSVT outcomes include higher SPLs and changes in laryngeal measures toward greater glottal efficiency and improved voice quality subsequent to therapy. The present data do not speak to loudness effects in these domains or, indeed, to the overall efficacy of LSVT. Our specific conclusion, based on the results obtained here, is simply this: Claims that louder speech leads to a general expansion of the vowel space, particularly with regard to F2, do not appear to be justified, at least for naturalistically elicited loud speech in typical speakers. This places limits on the degree to which one can think of loudness as a "natural" scaling method for all aspects of articulation and invoke such a mechanism to explain changes subsequent to loudness-based therapy. Tjaden and Wilding (2004) have pointed out that their loudness manipulations, obtained in single-session protocols, may differ from the long-term, intensive intervention involved in LSVT (see also Sapir et al., 2007, on possible differences between stimulated and trained loud speech). In other words, LSVT training may introduce a form of loud(er) speech distinct from what is obtained under more naturalistic conditions. This could represent one more example of how variation in tasks and instructions can lead to different changes in speech production behavior (cf. Garnier, Henrich, & Dubois, 2010; Lam & Tjaden, 2013; Lam, Tjaden, & Wilding, 2012).

Caveats

Most work on loudness has involved speakers of English, not German. It is an empirical question whether English speakers would show patterns comparable to those observed here. We suspect, however, that loudness effects in these two languages may be similar given that both have rather large vowel inventories, with multiple height levels and tense–lax variation. One notable difference between the two languages is that German includes front-rounded vowels, which could have the effect of restricting F2 variation, and F2 variation observed here was rather minor overall. However, the literature has not demonstrated strong or consistent loudness-related changes for F2 in English either.

It could be that languages with smaller vowel inventories (e.g., the common five-vowel system seen in Spanish, Russian, and Greek, among others) would allow different patterns of loudness-related formant variation. In fact, it is

[6]In the literature on typical speakers, authors have generally referred to loudness rather than effort, presumably because the former is objectively measurable; one exception is Traunmüller and Eriksson (2000).

striking that the literature on loud speech has, thus far, been dominated by languages with large vowel systems (viz., English, French, and Swedish). Future work is needed to evaluate language differences in this regard.

It is also possible that results for monolingual German speakers would be somewhat different from what we observed here, but this would be rather difficult to test directly given that virtually all young people in contemporary Germany learn one or more second languages. It was the case that the recording session was carried out almost entirely in German (with only a few occasional comments exchanged between some participants and the first author at the very beginning of the session; subsequently, participants interacted with the second author, in German). Our methods thus followed practices often adopted in studies of bilingual speakers, where conducting the research session in the target language should help participants minimize influences of other languages (e.g., Soares & Grosjean, 1984).

Although we endeavored to include a range of speech tasks, varying in the degree of coarticulatory influence and the extent to which speakers had to generate their own linguistic content, it is clear that considerably more work is needed to understand how speech production differs across speech tasks, whether task variation involves the complexity of phonetic and linguistic structure, or the manner in which participants are asked to speak.

Conclusions

This work evaluated formant differences in loud speech across speaking tasks and speakers, considering multiple vowels. The results showed effects of all three of these variables and lead to the conclusion that increasing loudness alone does not necessarily yield broad changes in vowel formants, considering both F1 and F2 variation, vowel height, and tense–lax differences. In this respect, our results recall the results of Schulman (1989) and Tasko and McClean (2004) and indicate that increasing loudness does not seem to represent a simple scaling mechanism, at least as far as vowels are concerned. Rather, effects appear to be specific to vowel, speaker, and task. As such, we conclude that there are multiple strategies whereby speakers achieve louder speech and formant changes should not be considered as a necessary concomitant of louder speech. Future work is needed to clarify the range of differences across various types of speaking tasks and establish the degree to which these results are cross-linguistically generalizable.

Acknowledgments

This work was supported by Ministry for Education Research Grant 01UG1411 and by the Leibniz Association. We express our thanks to the following: (a) our speakers, (b) Jörg Dreyer for technical assistance, and (c) Meghan Robitaille (supported by an Adelphi University Graduate Assistantship) for assistance with data processing.

References

Bates, D., Mäechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. https://doi.org/10.18637/jss.v067.i01

Baumgartner, C. A., Sapir, S., & Ramig, L. O. (2001). Voice quality changes following phonatory-respiratory effort treatment (LSVT) versus respiratory effort treatment for individuals with Parkinson disease. *Journal of Voice, 15,* 105–114.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/

Bond, Z. S., Moore, T. J., & Gable, B. (1989). Acoustic–phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *The Journal of the Acoustical Society of America, 85,* 907–912.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic–phonetic talker characteristics. *Speech Communication, 20,* 255–272.

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics, 24,* 263–282.

Darling, M., & Huber, J. E. (2011). Changes to articulatory kinematics in response to loudness cues in individuals with Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 54,* 1247–1259.

Dromey, C., & Ramig, L. O. (1998). Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation. *Journal of Speech, Language, and Hearing Research, 41,* 1003–1018.

Dromey, C., Ramig, L. O., & Johnson, A. B. (1995). Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: A case study. *Journal of Speech and Hearing Research, 38,* 751–764.

Fant, G. (1971). *Acoustic theory of speech production*. The Hague, the Netherlands: Mouton.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50,* 1241–1255.

Finnegan, E., Luschei, E. S., & Hoffman, H. T. (2000). Modulations in respiratory and laryngeal activity associated with changes in vocal intensity during speech. *Journal of Speech, Language, and Hearing Research, 43,* 934–950.

Fox, C. M., Morrison, C. E., Ramig, L. O., & Sapir, S. (2002). Current perspectives on the Lee Silverman Voice Treatment (LSVT) for individuals with idiopathic Parkinson disease. *American Journal of Speech-Language Pathology, 11,* 111–123.

Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research, 53,* 588–608.

Gilbert, R. C., Victor, N., Chandrasekaran, B., & Smiljanic, R. (2013). Intelligibility of speaking styles elicited by various instructions. *The Journal of the Acoustical Society of America, 134,* 4244.

Hazan, V., & Markham, D. (2004). Acoustic–phonetic correlates of talker intelligibility for adults and children. *The Journal of the Acoustical Society of America, 116,* 3108–3118.

Huber, J. E. (2007). Effect of cues to increase sound pressure level on respiratory kinematic patterns during connected speech. *Journal of Speech, Language, and Hearing Research, 50,* 621–634.

Huber, J. E., & Chandrasekaran, B. (2006). Effects of increasing sound pressure level on lip and jaw movement parameters and

consistency in young adults. *Journal of Speech, Language, and Hearing Research, 49,* 1368–1379.
Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., & Johnson, K. (1999). Formants of children, women, and men: The effects of vocal intensity variation. *The Journal of the Acoustical Society of America, 106,* 1532–1542.
Junqua, J.-C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America, 93,* 510–524.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82,* 1–26. https://doi.org/10.18637/jss.v082.i13
Ladefoged, P., & McKinney, N. (1963). Loudness, sound pressure, and subglottal pressure in speech. *The Journal of the Acoustical Society of America, 35,* 454–460.
Lam, J., & Tjaden, K. (2013). Intelligibility of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research, 56,* 1429–1440.
Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research, 55,* 1807–1821.
Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research, 14,* 677–709.
Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.
Lieberman, P., Knudson, R., & Mead, J. (1969). Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation. *The Journal of the Acoustical Society of America, 45,* 1537–1543.
Liénard, J.-S., & Di Benedetto, M.-G. (1999). Effect of vocal effort on spectral properties of vowels. *The Journal of the Acoustical Society of America, 106,* 411–422.
Lindblom, B. E. F., & Sundberg, J. E. F. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America, 50,* 1166–1179.
Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America, 124,* 3261–3275.
Matheron, D., Stathopoulos, E. T., Huber, J. E., & Sussman, J. E. (2017). Laryngeal aerodynamics in healthy older adults and adults with Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 60,* 507–524.
Mefferd, A. S., & Green, J. R. (2010). Articulatory-to-acoustic relations in response to speaking rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research, 53,* 1206–1219.
Picheney, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28,* 96–103.
Picheney, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29,* 434–446.
R Core Team. (2017). *R: A language and environment for statistical computing*. Retrieved from http://www.R-project.org/
Ramig, L. O., Bonitati, C. M., Lemke, J. H., & Horii, Y. (1994). Voice treatment for patients with Parkinson's disease: Development of an approach and preliminary efficacy data. *Journal of Medical Speech-Language Pathology, 2,* 191–209.
Ramig, L. O., Countryman, S., Thompson, L. L., & Horii, Y. (1995). Comparison of two forms of intensive speech treatment for Parkinson disease. *Journal of Speech and Hearing Research, 38,* 1232–1251.
Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M., & Thompson, L. L. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: A 2 year follow up. *Journal of Neurology, Neurosurgery, & Psychiatry, 71,* 493–498.
Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *The Journal of the Acoustical Society of America, 125,* 2288–2298.
Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language, and Hearing Research, 53,* 114–125.
Sapir, S., Spielman, J. L., Ramig, L. O., Hinds, S. L., Countryman, S., Fox, C., & Story, B. (2003). Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on ataxic dysarthria: A case study. *American Journal of Speech-Language Pathology, 12,* 387–399.
Sapir, S., Spielman, J. L., Ramig, L. O., Story, B. H., & Fox, C. (2007). Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research, 50,* 899–912.
Schmidt, C. P., Gelfer, M. P., & Andrews, M. L. (1990). Intensity range as a function of task and training. *Journal of Voice, 4,* 30–36.
Schulman, R. (1985). Dynamic and perceptual constraints of loud speech. *The Journal of the Acoustical Society of America, 78,* S37.
Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America, 85,* 295–312.
Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America, 118,* 1677–1688.
Soares, C., & Grosjean, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory & Cognition, 12,* 380–386.
Stevens, K. N., & House, A. S. (1955). Development of a quantitative description of vowel articulation. *The Journal of the Acoustical Society of America, 27,* 484–493.
Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America, 122,* 1111–1129.
Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America, 84,* 917–928.
Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America, 79,* 1086–1100.
Tartter, V. C., Gomes, H., & Litwin, E. (1993). Some acoustic effects of listening to noise on speech production. *The Journal of the Acoustical Society of America, 94,* 2437–2440.
Tasko, S. M., & McClean, M. D. (2004). Variations in articulatory movement with changes in speech task. *Journal of Speech, Language, and Hearing Research, 47,* 85–100.
Tjaden, K., Lam, J., & Wilding, G. E. (2013). Vowel acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions. *Journal of Speech, Language, and Hearing Research, 56,* 1485–1502.
Tjaden, K., Sussman, J. E., & Wilding, G. E. (2014). Impact of clear, loud, and slow speech on scaled intelligibility and speech

severity in Parkinson's disease and multiple sclerosis. *Journal of Speech, Language, and Hearing Research, 57,* 779–792.

Tjaden, K., & Weismer, G. (1998). Speaking-rate-induced variability in F2 trajectories. *Journal of Speech, Language, and Hearing Research, 41,* 976–989.

Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research, 47,* 766–783.

Tjaden, K., & Wilding, G. E. (2005). Effect of rate reduction and increased loudness on acoustic measures of anticipatory coarticulation in multiple sclerosis and Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 48,* 261–277.

Traunmüller, H., & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America, 107,* 3438–3451.

Weirich, M., & Simpson, A. (2013). Investigating the relationship between average speaker fundamental frequency and acoustic vowel space size. *The Journal of the Acoustical Society of America, 134,* 2965–2974.

Wenke, R. J., Cornwell, P., & Theodoros, D. G. (2010). Changes to articulation following LSVT and traditional dysarthria therapy in non-progressive dysarthria. *International Journal of Speech-Language Pathology, 12,* 203–220.

Whitfield, J. A., Dromey, C., & Palmer, P. (2018). Examining acousic and kinematic measures of articulatory working space: Effects of speech intensity. *Journal of Speech, Language, and Hearing Research, 61,* 1104–1117.