Self-perception of vocal and articulatory effort in consonant production by native Swedish speakers

Iris Gordon-Bouvier[1], *Josefine Kyhle*[1], *Anita McAllister*[1], *Hanna Norman*[1], *Sarah Paues*[1], *Camille Robieux*[2] *and Sofia Strömbergsson*[1]
[1]*Karolinska Institute, Stockholm, Sweden*
[2]*University of Marseille, France*

Abstract

This study investigates native Swedish speakers' self-perception of vocal and articulatory effort in a speech sound production task, in an attempt to identify patterns amongst Swedish speakers as well as to make cross-language comparisons with previously reported data for French speakers. Vocal and articulatory effort refers to the level of exertion required to produce a sound. Previous research has primarily focused on physiological rather than self-perceived measurement.

In a partial replication of a French study on vocal and articulatory parameters, twenty-two healthy native speakers of Swedish aged 21-50 were presented with 220 pairs of nonsense syllables that they were asked to produce aloud. Each pair included up to three contrasts between the consonants in terms of context, voicing, and place and manner of articulation. Some subjects were instructed to select the item from each pair that they felt was easier to produce, while others were instructed to choose the one they perceived as more difficult.

The results indicate that subjects perceived voiced consonants as requiring a greater level of effort than voiceless consonants. A higher rate of self-perceived effort was also reported for unvocalic (isolated) consonants compared with intervocalic consonants, regardless of voicing.

These findings may be significant for professionals in the field of speech therapy, by providing a baseline for structuring clinical voice therapy based on subjective experience.

Introduction

During speech, the glottis varies from wide open to tightly shut, due to variations in the level of air pressure produced by the lungs and laryngeal muscle activity (Johnson, 2012).

When air particles travel through a narrow passage such as the glottis, the particles accelerate and thin out, lowering air pressure in the passage. Due to the elasticity of the glottis, the low air pressure causes the vocal folds to contract and the airflow is interrupted. The passage remains closed until the subglottic pressure is built up to the point where it pushes the glottis open again, and the cycle repeats over and over. Voicing, or phonation, occurs when the vocal folds vibrate in this manner, according to the myoelastic-aerodynamic theory of voice production (Van den Berg, 1958).

Vocal fold vibration is the usual source of sound in vowels, and our vocal tract serves as an acoustic filter that modifies the sound (Johnson, 2012). This model is known as the source-filter theory of speech production (Fant, 1960).

Consonants are sounds that are produced through a narrowing or complete closure anywhere between the glottis and the lips, using the active articulators, i.e. the parts of the vocal tract that can be manipulated at will (Engstrand, 2004). Depending on whether or not the vocal folds are engaged, consonants are either voiced or voiceless.

Vocal and articulatory effort refers to the level of exertion required to produce a sound. Many patients with voice disorders, such as dysphonia, nodules or polyps, report a general increase in vocal effort (Rosenthal et al, 2014).

In a clinical situation, a speech therapist can recommend and train patients in various vocal

exercises, with thoroughly documented effects (ASHA, 2005). However, beyond patients' informal reports of improved vocal function, these effects are generally based on physiological measurements of human vocal effort rather than scientific studies of self-perception in healthy subjects. In a study, focusing on laryngeal resistance (LR, subglottic pressure divided by average airflow through the glottis) a correlation was found between LR and a trained professional's auditory perception of participants' vocal quality (Grillo et al., 2009). To date, limited research has been done into the physiological parameters that contribute to the perception of vocal and articulatory effort. The present study aimed to understand how Swedish healthy subjects perceive vocal and articulatory effort in a production task, and whether the findings differ from those made in a preliminary study performed on French speaking subjects (Robieux, 2015). Robieux's preliminary results indicate that native speakers of French perceive that a significantly greater effort is needed to produce voiced consonants than voiceless regardless of context. The two main research questions of the present study were whether there would be corresponding patterns in the self-perception of Swedish speech sound production, and to investigate differences and similarities between Swedish and French results.

There are a number of articulatory differences between Swedish and French. One of these is aspiration. In Swedish, voiceless plosives are usually aspirated when they occur in an initial position within a stressed syllable, unless preceded by /s/ (Engstrand, 1999). In verbal communication prior to the study, Robieux expressed an expectation that aspiration on voiceless plosives could affect perceived vocal and articulatory effort to the extent that Swedish subjects might rate voiceless consonants as more difficult to produce than voiced. As native speakers of Swedish, the authors of the present study proposed a somewhat different hypothesis, namely that the differences between the two languages would not be enough to override the general increase in effort required for phonation. Voiced consonants involve a high level of coordination between subglottic pressure and the vocal folds, together with a constriction or complete closure of the vocal tract (Ohala, 1983). Furthermore, these changes occur extremely fast, in a matter of milliseconds. The expectation was that the results would reflect this increased level of activity. A secondary hypothesis was that unvocalic (isolated) consonants would be perceived as more effortful than intervocalic consonants (uttered between two vowels), due to the need to coordinate high air pressure, vocal articulators and, in the case of voiced consonants, phonation, in a very short space of time.

The present study may provide valuable insight that can be exploited for the purpose of improving voice therapy. Identification of those speech sounds perceived by healthy subjects as requiring more effort to produce can help caregivers to adapt treatment by reducing the sense of vocal effort while still effectively treating voice disorders.

Method

Subjects

A convenience sampling of twenty-two adult subjects was carried out, with thirteen women and nine men ranging in age from 21-50 years. All subjects were native speakers of Swedish, with no history of reading, speech or voice disorders.

Stimuli

Each subject was presented with 220 pairs of items made up of nine different consonants, both voiced and voiceless - /b/, /p/, /d/, /t/, /g/, /k/, /v/, /f/, and /s/ - combined with the vowel /a/. The consonants contrasted in terms of place and manner of articulation, and comprised labial, dental and velar plosives and fricatives. Consonants were presented in four different contexts: nonsyllabic (unvocalic #C#), monosyllabic (prevocalic #CV, postvocalic VC#) and disyllabic (intervocalic VCV). Pairs were presented randomly, each pair occurring twice in reverse order; item 1-item 2 versus item 2-item 1. There were two different test forms; one instructing subjects to focus on the sounds that felt easier to produce, and the other instructing subjects to select the more difficult option. Both tests contained the same pairs of items, but in a different order.

Procedure

The experiment was of a between-subjects design, and was carried out in a quiet, familiar setting. Subjects were given both verbal and written instructions, and were asked to sign a

consent form stating that they agreed to their data being used in the study.

Prior to starting the test, participants were presented with a list of speech sounds for training, enabling experiment leaders to correct pronunciation if necessary, for example if a voiceless consonant was pronounced in a voiced manner. Such corrections were also made during the actual test.

The task consisted of reading each pair aloud using a normal speaking voice, with a brief pause between the two items, and to circle one item per pair, depending on specific instructions. Repetition of each item was permitted as necessary, and participants were instructed to leave no blanks. Answers were recorded by hand, and no digital recordings were made. The test was completed in a single session, with a mean time of approximately 5-8 minutes per page, although there was no time limit.

Data were entered manually in an Excel spreadsheet, with one data sheet per subject, and answers were converted to the dichotomous values 1 and 2. χ^2 analyses were then performed on the relevant parameters using the SPSS software program.

Results

Two primary contrasts were selected and analysed; voiced-voiceless and unvocalic-intervocalic.

Combined responses from all participants rendered 704 instances where the main contrast was whether consonants were voiced or voiceless. In 407 (57.8 %) of these cases, subjects chose the voiced consonant as requiring more vocal effort (see *figure 1).*

There were 396 total instances where the main contrast between the consonants was unvocalic (#C#) versus intervocalic (VCV). In 259 instances (65.4 %), the unvocalic consonant was selected as requiring more effort to produce *(*see *figure 2)*. A χ^2 analysis of these primary contrasts yielded a significant result ($p<0.001$) for both parameters. For the contrast voiced-voiceless χ^2 (1, N=704) = 17.19, $p<0.0001$. For the contrast unvocalic-intervocalic χ^2(1, N=396) = 37.7, $p<0.0001$.

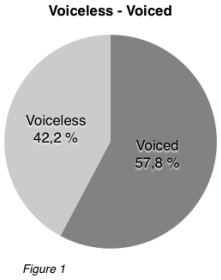

Figure 1

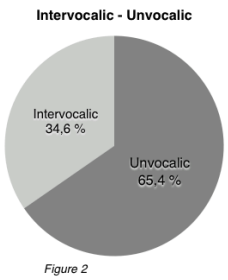

Figure 2

Figure 1 shows how test subjects rated vocal effort required to produce consonants where the main contrast was voiced-voiceless, while the main contrast in figure 2 was unvocalic-intervocalic. Both figures show how often subjects rated the parameters as requiring more effort.

Discussion

The two questions that were posed within the present study were whether a significant pattern would be found in Swedish speakers' self-perceived vocal and articulatory effort in producing consonants, and how this would compare to data from the French study (Robieux, 2015).

The main hypothesis underlying this study was that Swedish speakers would rate voiced consonants as being more difficult to produce than voiceless consonants, hence that our result would match those rendered by Robieux. A secondary hypothesis was proposed, namely that unvocalic (isolated) consonants would be perceived as requiring more vocal effort than intervocalic consonants.

Indeed, analysis of data yielded a significant result, in alignment with both the primary and secondary hypotheses. In the majority of cases where subjects were asked to choose between voiced and voiceless consonants, voiced were rated as being more difficult than unvoiced, corroborating Robieux's findings. In the majority of the cases where the main contrast was unvocalic-intervocalic, unvocalic consonants were deemed harder to produce.

Interpretation of data

The production of a plosive requires a complete closure of the vocal tract, resulting in a build-up of pressure, which is subsequently released when producing the sound. If the plosive is voiced, the vocal folds are activated before or upon the burst. This subglottal pressure build-up and release may require more effort, also involving the coordination of the speaker's

vocal folds and vocal tract when the plosive is voiced (/b/, /d/, /g/). If the plosive is voiceless, (/p/, /t/, /k/), the effort may be smaller.

When producing an open vowel, pressure from the lungs is constant, creating a flow through the vibrating vocal folds. If a plosive is produced in an intervocalic context, the production of the plosive may be perceived by the speaker as "riding the wave" of this subglottic pressure thus appearing to require less effort. Furthermore, there is some evidence that in a VCV context, where the consonant is a plosive, the movement of the articulators occur almost simultaneously (Gay, 1977). This could also explain why subjects rated intervocalic consonants as requiring less vocal effort.

It could be argued that isolated consonants rarely occur in spontaneous speech in Swedish. Therefore they may be less automated than sequences combining them with vowels would be. However, in traditional voice therapy exercises using isolated consonants are not uncommon (Carlsson et al., 1985).

Possible factors affecting outcomes

The test material was originally designed for French speakers. No instruction was given to the Swedish speakers with regards to pronunciation or duration of the open vowel sound. In Swedish, the vowel sounds /a/ and /ɑ/ are both possible, a fact that may have affected the results (Handbook of IPA, 1999). The rounding of the lips necessary to produce the Swedish vowel /ɑ/ may result in a slight increase in perceived effort compared to /a/, as it requires activation of the orbicularis oris muscles around the mouth (Engstrand, 2004). This freedom to interpret the vowel sound could lead to unexpected differences not only between Swedish and French test results, but also within the Swedish study, possibly even within individual subjects. Had instructions on the pronunciation of the vowel been included in the pronunciation practice sheet, potential pitfalls such as these might have been reduced.

Implications for future studies

The contrasts explored in the present study are by no means exhaustive. An example of a parameter that would be interesting to investigate in Swedish is the aforementioned aspiration with regard to voiceless plosives. It would be worthwhile to explore this particular contrast to see whether the presence or absence of aspiration affects perceived vocal and articulatory effort, particularly in light of Robieux's hypothesis regarding Swedish. However, the test material used in the present study was developed for native French speakers and thus not designed with this specific contrast in mind. Prior to the start of the experiment, speech sounds present in French and not in Swedish had been excluded from the test forms developed by Robieux. However, no items were added to accommodate for Swedish speech sounds. Furthermore, analysis of data for the voiced-voiceless contrast made no distinction between plosives and fricatives. It is therefore not appropriate to draw conclusions about aspiration within the limits of the present study. The contrast aspirated-unaspirated voiceless plosives would be a valuable addition to a future test on Swedish subjects.

The present study is still at an early stage of development, yet opens up possibilities for additional research on vocal effort, and how voice therapy could be adapted according to subjective experiences. It is worth stating that a high level of perceived vocal and articulatory effort is not necessarily something to avoid during voice therapy. Rather, the findings offer some insight into how to structure therapy according to a patient's personal needs.

The general format lends itself to replication for other languages. The test could be adapted for native speakers of other languages and results of these languages correlated. Documented similarities between results for different languages could further encourage voice therapists to share ideas internationally in order to increase their knowledge and range of therapeutic exercises.

A range of information was gathered on participants, including gender, education level, presence of asthma, history of smoking, and further languages spoken. Another parameter that was not taken into account was participants' awareness of vocal function and their own voice. It would be interesting and worth exploring possible correlations between some of these variables and self-perception of vocal effort.

References

Carlsson, I, Eriksson, C, Persson, E L, Svanberg, P (1985). *Rösthantverket - en idébank för kliniskt erfarna röstlogopeder*. (Dissertation) Karolinska Institute, institution of phoniatrics and speech pathology, Huddinge sjukhus, Sweden. (Swedish)

Engstrand, O (2004). *Fonetikens grunder*. Sweden: Studentlitteratur, 121. (Swedish)

Engstrand, O (1999). Illustrations of the IPA: Swedish, in *Handbook of the International Phonetic Association*. United Kingdom: Cambridge University Press, 140-142.

Fant, G (1960). *Acoustic theory of speech production.* Netherlands: Mouton Publishers, 15-17.

Gay, T (1977). Articulatory movements in VCV sequences. *Journal of the Acoustic Society of America*, 62, 183-193.

Grillo, E U, Perta, K, & Smith, L (2009). Laryngeal resistance distinguished pressed, normal, and breathy voice in vocally untrained females. *Logopedics Phonatrics Vocology*, *34*(1), 43-48.

Johnson, K (2011). *Acoustic and auditory phonetics*. Chichester: Wiley-Blackwell, 25, 156.

Ohala, J (1983). The origin of sound patterns in vocal tract constraints. *The production of speech,* 189-216.

Rosenthal, A L, Lowell, S Y, & Colton, R. H. (2014). Aerodynamic and acoustic features of vocal effort. *Journal of Voice*, *28*(2), 144-153.

Robieux, C (2015). Effect of voicing on the self-perception of effort in French consonants production. (Accepted for publication in *Proceedings of the ICPhS 2015*, Glasgow, UK).

Van den Berg, J (1958). Myoelastic-aerodynamic theory of voice production, *Journal of Speech and Hearing Research* 3(1): 227-244.

ASHA (American speech-language-hearing association) (2005). Compendium of EBP Guide-line and Systematic Reviews. Retrieved 25/3-15. www.asha.org/members/ebp/compendium/