Nonverbal Vocalisations – A Forensic Phonetic Perspective

Angelika Braun
University of Trier, Germany
brauna@uni-trier.de

Abstract

This contribution approaches nonverbal vocalisations from an angle which is probably quite different from most other perspectives – its usability for forensic speaker comparison purposes. Thus the question is whether and if so, to what extent, nonverbal vocalisations are speaker specific. In this paper, it is argued that it is not so much any one individual trait which is speaker specific but rather a behavioral pattern consisting of various elements. How these vocalisations are covered in forensic phonetic reports is described. Various aspects of the behavioral pattern are dealt with: hesitations/filled pauses, breathing, clicks, question tags, tempo, and laughter.

1 Forensic Voice Comparison

The comparison of a recorded voice sample of an unknown speaker (the perpetrator) with that of a known speaker (the suspect) is one of the classical forensic phonetic tasks. The most widely used methodological approach to this task is the auditory acoustic one, i.e. an approach which combines an in-depth auditory phonetic analysis with acoustic measurements (Gold and French 2011).

One of the principal problems to be overcome in this work is that speech – unlike fingerprints or DNA – forms part of human behavior which may well be subject to short-term or long-term variability, rather than being an invariable anatomical or physiological trait. Short-term variability of the human voice includes changes with emotion (Braun and Heilmann 2012), health status (Baken 1987), or even time of day (Garrett and Healey 1987); long-term variability includes changes with age (Linville 2001) or with moving to a different part of the country and adopting some regional characteristics of that area (Kiesewalter 2019) etc.

Wolf (1972) and Nolan (1983: 11) have defined sets of requirements for the ideal parameter to be used in forensic voice comparison:

-availability even in small amounts of material
- robustness to disguise
- low intraspeaker variability
- high interspeaker variability
- measurability
- durability, i.e. remaining unchanged over time.

Traditionally, the criteria which have been used for decades fall into three categories: voice, speech, and manner of speaking. (Künzel 1987, Wagner 2019).

Features of voice encompass mean fundamental frequency as well as its variability and baseline (*Lösungstiefe* in German). Features of speech contain information on regional, social and individual pronunciation patterns etc. They also include e.g. speech disorders and mispronunciations of sounds in general. Verbal mannerisms (e.g. frequent use of "like" or "so-called") and question tags, which are highly variable in German, will also fall into this category. Manner of speaking includes various nonverbal features such as hesitations, question tags, speaking tempo, pausing, breathing patterns, laughter, and clicking sounds made by, e.g., ill-fitting dentures (cf. Künzel 1987 for an early account; Jessen 2008).

Since speaking is – to a large extent – behavior, no single one of these parameters can be expected to constitute speaker specificity. Instead, all of them contribute to a behavioral pattern that establishes speaker individuality.

Nonetheless, it is worthwhile to study to what extent the various parameters mentioned can be shown to be speaker specific. Hence, forensic

phoneticians have been looking into this issue, studying hesitations, tempo, pausing, and breathing patterns. Below, some of the findings is summarized, and ongoing research is described.

2 Research Issues

The basic question behind all the research reported here is which, if any, nonverbal cues can be shown to be speaker-specific to an extent which makes them usable in forensic voice comparison.

3 Hesitations

Hesitations form part of a whole behavioral pattern, which consists of pausing, hesitations, repetitions, and false starts as evidence of speech planning and possibly laughter in addition. With the forensic perspective in mind, an integrated view on all of these aspects of hesitation behavior should be adopted (cf. McDougall et al. 2019).

3.1 A Pilot Study on German

In a pilot study the phonetic distribution of hesitation markers of eight speakers of German was addressed (Braun and Rosin 2015).

Speakers were recorded (spontaneous speech on a controversial political issue) on three different occasions with a few days' time between recordings. The total recording per person time varied between 44 and 75 minutes.

The following vocal manifestations of hesitation were studied: *uh* (inserted vowel), *uhm* (inserted vowel plus nasal consonant), *mh* (nasal), FVL (final vowel lengthening) *ich habe* [ɪç haːbəː], FCL (final consonant lengthening) *ich muss* [ɪç mʊsː] ICL (initial consonant lengthening) *so* [zːoː], and IVL (initial vowel lengthening, cf. e.g. *und äh* [ðːntɛː].

Speakers were fairly consistent in their hesitation behavior across sessions (cf. Figure 1). However, the number of hesitations per minute was not very speaker specific: speakers 2 and 6 show exactly the same average; the numbers for speakers 1, 7, and 10 are practically identical as well (Braun and Rosin 2015). That is why we need a closer look at the distribution of the different hesitation markers.

The results demonstrate the following: First, the distributions of the various hesitation markers are very similar across sessions per speaker. Secondly, those subjects who exhibit the same frequency of hesitation use differ considerably in their preferred hesitation sounds. It is thus easy to tell the speakers apart.

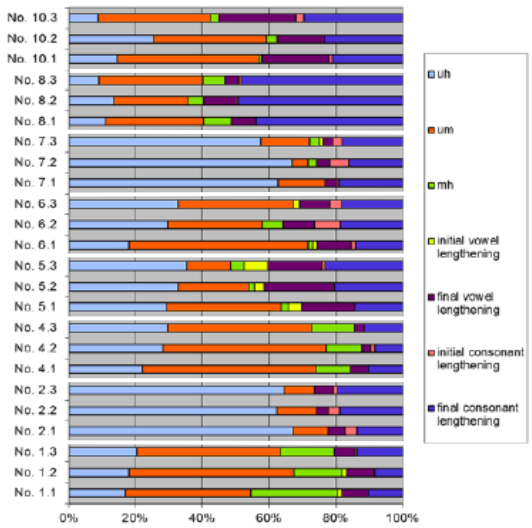

Figure 1. Cumulative distribution of the various hesitation types (German). Numbers on the left indicate subjects and session.

3.2 A Pilot Study on Spanish

A similar study was conducted with Spanish speakers. The results are shown in Figure 2.

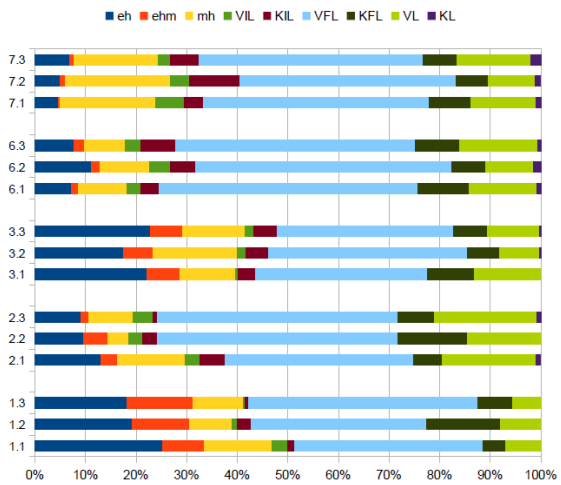

Figure 2. Cumulative distribution of the various hesitation types (Spanish). Numbers on the left indicate subjects and session.

The consistency across sessions is similar to that of the German speakers, but the preferred hesitation marker is different: the Spanish speakers all use final vowel lengthening most frequently. They differ in their second choice of hesitation marker, though. This result is very different from that of the German speakers, who showed preferences for different hesitation markers, none

of them final lengthening. These results indicate that the patterns may be language-specific as well as individual. Further steps will focus on adding more subjects to the database and including other aspects of hesitation behavior.

4 SpeakingTempo

Speaking tempo can be defined as the number of linguistic units per time unit. There are various ways of measuring it. Linguistic units may be words, syllables, or sounds. Time units may be minutes or seconds. There is a number of reasons why it does not seem particularly advisable to use the word as the linguistic unit to be measured. Especially in languages like German which make abundant use of compounds sometimes leading to monstrous words like *Donaudampfschifffahrts-kapitänswitwe* 'the widow of a Danube steamer's captain', word length is so variable that it is not a measure which can easily be compared between speakers. It also has to be taken into account that forensic speech samples are quite short most of the time, which lowers the probability of working with a representative distribution of word length. On the other hand, choosing the sound as a unit to be measured may prove to be highly impractical because it is very tedious for longer texts. This leaves one with the syllable as the unit of choice – at least in the forensic context. In relation to the syllable, two further questions arise:

(a) Should syllable rate or articulation rate be measured?

(b) Should linguistic or phonetic syllables be measured?

The difference between syllable and articulation rates consists in the treatment of pauses: whereas syllable rate is calculated by dividing the number of syllables by the duration of the utterance in seconds, articulation rate is defined as the number of syllables of net speech per second, i.e., after the pauses have been removed. Thus, the effect of overall communicative setting and the degree of enunciation which the various settings call for is eliminated to some extent.

With respect to the type of syllable, there is a choice between phonetic and linguistic syllables. There are good reasons for studying either, but it has to be kept in mind that they address different aspects of tempo: The number of linguistic (canonical) syllables per second does not so much reflect the actual velocity of the articulator moves but represents a crude overall measure of tempo. By measuring phonetic syllables we gain insight into reduction processes generated by the individual speaker in informal communication. An example would be the sentence *Wir haben heute keine Zeit*. 'We don't have time today'. In its canonical form, the utterance consists of eight syllables in German, but it may also be realized as *Wir ham heut kein Zeit*, which consists of only five syllables. In the forensic context, it is important to consider the actual movements of the articulators, which makes articulation rate the parameter of choice (cf. Schilz 2008). Furthermore, it may be useful to calculate the difference between the number of linguistic and phonetic syllables, because it represents the degree of reduction which a given speaker decides to use. Thus, analogous to the *Lautzahlminderungs-quotient* 'sound reduction quotient' introduced by Hildebrandt (1962), a "syllable reduction quotient" (SRQ) could be defined as

$$\text{SRQ} = 10 - \frac{10\ x\ number\ of\ phonetic\ syllables}{number\ of\ linguistic\ syllables}$$

A positive number represents a reduction in the number of syllables which are actually produced as compared to those that ought to be produced, whereas a negative number means that there are epenthetic syllables (e.g. *Halt Dein Mauel* 'shut up'). The amount of syllable reduction in comparable communicative settings could prove to be a valid parameter in speaker comparison. It has to be noted, though, that SRQ does not capture reduction processes such as plosive lenitions or *r*-vocalisations, which do not change the number of syllables.

5 Breathing patterns

Evidence from previous research demonstrates that breathing patterns meet important criteria for voice comparisons: they are individual, and they remain constant throughout life (Benchetrit et al. 1989, Eisele et al. 1992). The question is whether this individuality also applies to speech breathing and whether it can be derived from the typical forensic recording, which is short and telephone-transmitted. A preliminary study on the usability of this type of recording for analyzing breathing patters based on 150 speakers of the telephone-transmitted DIGS corpus showed that breathing was audible in 79% of cases and formants could be measured in 67% of the recordings (Schwerdt

2019). Kienast and Glitza (2003), when studying ten speakers, found promising aspects of individuality in speech breathing, e.g. concerning the frequency of breathing cycles, the distribution of breathing types (oral, nasal, combined) as well as the spectral composition of the breathing noises. This is currently being followed up in a project at the University of Trier.

6 Click Sounds

In rare cases, "extraneous" noises are encountered in forensic recordings which originate for example from ill-fitting dentures. They reflect sucking noises generated by dentures coming apart from the palate. In an actual case, this proved to be significant because it occurred in a young speaker below 30 years of age. It turned out that the suspect had an underbite for which he had to wear dentures that did not fit very well. This agreement between questioned and reference recordings was considered to be rare and played a major role in the probability rating in the forensic report.

A second source of clicking noises in forensic materials exists which can be described as a "dry-mouth-syndrome". Either owing to certain types of medication (e.g. Beta-blockers for high blood pressure) or as a consequence of situational stress, the sucking sound of the tongue disengaging from the palate may be clearly audible and measurable. It should be noted that these clicking sounds are different from the ingressive velaric clicks studied by Gold et al. (2013) who found those sounds to be of little discriminative value.

On the other hand, excessive salivation may lead to frequent swallowing which is also taken into account in forensic reports[1].

7 Laughter

Laughter is only rarely encountered in the forensic setting, but if it does occur, it needs to be included in the report (cf. Hirson 1995 for a rare account of an actual case). An example from a rip-deal case is given in Figure 3 below: The perpetrator laughs at the victim as part of the strategy to persuade him to "invest" a large sum of money.

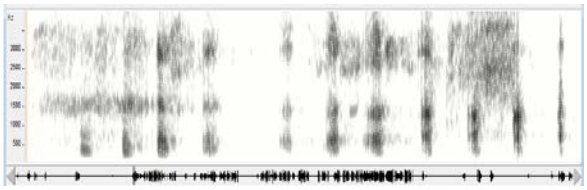

Figure 3. Laughter in a forensic case. (Frequencies are displayed up to 4 kHz due to telephone transmission.)

This case awaits a reference recording from the suspect. In analyzing laughter, the framework developed by Trouvain (2014) is very useful.

8 Question tags

Strictly speaking, question tags are not nonverbal vocalisations. However, just as laughter or hesitations, they do not represent lexical meaning and their use is highly individual. From a functional (and certainly from a forensic phonetic) perspective, they are comparable to hesitations. In German, a plethora of question tags can be found. They are in part regional and in part individual. In the Low German dialect area, one will primarily encounter *nich* [nɪç], *nich wahr* [nɪç vaː], and *ne* [nɛ] or [nə]. In the Berlin area, *wa*, [va] forms the most frequent question tag whereas it is *woll* [vɔł] in and around the city of Dortmund. In the middle and southern parts of Germany, a number of variants of *gell* are found. This may be *gell* [gɛl], *gelt* [gɛlt], *gelle* [gɛlə] or just *ge* [gɛ]. In the Southwest, *oder* is the tag commonly used (cf. *Atlas zur deutschen Alltagssprache* [round 2] for a map showing the regional distribution). For all of these, there are individual variants.

The frequency of occurrence, the lexical form, the prosodic as well as the formant structures of these tags are routinely looked at in forensic phonetic analysis.

9 Conclusion

Nonverbal vocalisations are already a well-established element in forensic phonetic analysis. However, there is a need for larger databases to be analyzed, for in-depth statistical analysis, and for an integrated approach which aims at behavioral patterns rather than individual features.

[1] This can be encountered in one of the German TV weather announcers (Donald Bäcker) who swallows frequently during his announcements.

10 References

Atlas zur deutschen Alltagssprache. 2003 ff. http://www.atlas-alltagssprache.de (last accessed on 20.02.2020).

Ronald J. Baken. 1987. *Clinical Measurement of Speech and Voice*. London: Taylor & Francis.

G. Benchetrit, S.A. Shea, T. Pham Dinh, S. Bodocco, P. Baconnier and A. Guz. 1989. Individuality of breathing patterns in adults assessed over time. *Respiration Physiology* 75: 199-210.

Angelika Braun. 2017. Forensische Sprach- und Signalverarbeitung. In: Bockemühl, Jan (Hg.): *Handbuch des Fachanwalts Strafrecht*. 7. Auflage. Köln: Wolters Kluwer. S. 1800-1823.

Angelika Braun and Christa M. Heilmann. 2012. *Synchron*Emotion, Frankfurt am Main: Peter Lang.

Angelika Braun and Annabelle Rosin. 2015. On the speaker specificity of hesitation markers – a pilot study. In: *Proceedings of XVIIIth International Congress of Phonetic Sciences*, Glasgow, 5p

J. H. Eisele, B. Wuyam, G. Savourey, J. Eterradossi, J. H. Bittel and G. Benchetrit. 1992. Individuality of breathing patterns during hypoxia and exercise. *Journal of Applied Physiology* 72, 2446-2453.

K. L. Garrett and E. Ch, Healey. 1987. An acoustical analysis of fluctuations in the voices of normal adult speakers across three times of day. *Journal of the Acoustical Society of America* 82: 58-62.

Erica Gold and Peter French. 2011. International practices in forensic speaker comparison, *The International Journal of Speech, Language and the Law* 26(1): 293–307.

Erica Gold, Peter French, and Philip Harrison. 2013. Clicking behavior as a possible speaker discriminant in English. *Journal of the International Phonetic Association* 43: 339-349.

Bruno Hildebrandt. 1961. Die arithmetische Bestimmung der durativen Funktion. Eine neue Methode der Lautdauerbewertung, *Zeitschrift für Sprachwissenschaft, Phonetik und Kommunikationsforschung* 14: 328-336.

Allen Hirson. 1995. Human laughter – A forensic phonetic perspective. In: Angelika Braun and Jens-Peter Koester (eds), *Studies in Forensic Phonetics*. Trier: Wissenschaftlicher Verlag. 77-86.

Michael Jessen. 2008. *Phonetische und linguistische Prinzipien des forensischen Stimmenvergleichs*. Lincom.

Miriam Kienast and Florian Glitza. 2003. Respiratory Sounds as an Idiosyncratic Feature in Speaker Recognition. In: *Proceedings of XVth International Congress of Phonetic Sciences*, Barcelona. 1607-1610.

Carolin Kiesewalter. 2019. *Zur subjektiven Dialektalität regiolektaler Aussprachemerkmale des Deutschen*. Stuttgart: Steiner.

Hermann Künzel. 1987. *Sprechererkennung. Grundzüge forensischer Sprachverarbeitung*. Heidelberg: Kriminalistik-Verlag.

Claudio L. Lafortuna, Alberto E. Minetti and Piero Morgnoni. 1984. Inspiratory flow patterns in humans. *Journal of Applied Physiology* XX: 1111-1119.

Sue Ellen Linville. 2001. *Vocal Aging*. San Diego: Singular.

Kirsty McDougall, Richard Rhodes, Martin Duckworth, Peter French and Christin Kirchhübel. 2019. Application of the 'Toffa' Framework to the Analysis of Disfluencies in Forensic Phonetic Casework. In: Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren (eds), *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia. Canberra, Australia: Australasian Speech Science and Technology Association Inc.,

Francis Nolan. 1983. *The phonetic bases of speaker recognition*. Cambridge: CUP.

Jessica Schilz. 2008. Wie individualtypisch ist die Artikulationsrate unter Berücksichtigung phonetischer bzw. linguistischer Silben? Eine empirisch vergleichende Studie. M.A. Thesis, University of Trier.

Jana Schwerdt. 2019. Atmung in der forensischen Phonetik. Eine Untersuchung zu hörbaren Atemgeräuschen. Term Paper, University of Trier.

Jürgen Trouvain. 2014. Laughing, Breathing, Clicking – The Prosody of Nonverbal Vocalisations. Proceedings of Speech Prosody,

Isolde Wagner. 2019. In: Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren (eds), *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia. Canberra, Australia: Australasian Speech Science and Technology Association Inc.,

Jared J. Wolf. 1972. Efficient acoustic parameters for speaker recognition. *Journal of the Acoustical Society of America* 51: 2044-2056.