

# PERCEPTION OF VOICE ONSET TIME BY AUSTRIAN AND GERMAN LISTENERS

Petra Hödl

Department of Linguistics, University of Graz  
petra.hoedl@edu.uni-graz.at

## ABSTRACT

In word-initial position, Standard German exhibits a two-way phonemic contrast between plain voiceless unaspirated stops /b d g/ and voiceless aspirated stops /p t k/. For Austrian German, it has however been claimed that there is a tendency to reduce aspiration. This paper deals with the potential perceptual consequences of this production tendency. A categorisation task was conducted where Austrian (n=33) and German listeners (n=47) had to label semi-manipulated speech stimuli varying in their voice onset time (VOT), place of articulation and burst/vowel quality. For each listener, psychometric curves and perceptual boundaries were calculated. While a comparison between the two listener groups did not reveal shifted perceptual boundaries overall, significant individual differences regarding the degree of categoricity of perception were found. Results also indicate that VOT/aspiration is not the only relevant voicing cue but that Austrian and German listeners are heavily influenced by the burst/vowel quality of the presented stimulus.

**Keywords:** Austrian German, voice onset time, speech perception, phoneme categorisation.

## 1. INTRODUCTION

In Standard German, the two-way phonemic contrast between “voiced” (/b d g/) and “unvoiced” (/p t k/) stops is word-initially realized as an aspiration contrast rather than a ‘true’ voicing contrast. /b d g/ are unaspirated and show a short-lag positive voice onset time (VOT) [10] while /p t k/ are aspirated and show a long-lag positive VOT. Standard German is therefore classified as an ‘aspirating’ language [2].

While this observation is true for the pronunciation of stops by North German speakers [7], speakers of some (Southern) varieties of German, such as Austrian German, are often said to show reduced aspiration [8, 12]. Especially in the case of /p/ and /t/ this tendency may even lead to a neutralization of the voicing contrast [8].

However, it is not clear how this tendency for reduced aspiration affects *perception*. While there is a considerable number of production studies on voicing and aspiration in different varieties of

German [4, 7, 11], perception studies are sparse. So far, there is no experimental data available on VOT perception by Austrian German listeners. It is also unclear how other phonetic features such as place of articulation or burst/vowel onset characteristics interact with VOT in these listeners. The aim of this study is hence to shed light on this topic.

During speech perception, listeners need to make more-or-less categorical decisions about the lexical identity of the speech input. These decisions could be impeded if the acoustic-phonetic cues in question lack linguistic relevance in the listener’s native language or if they are regarded as unreliable based on experience from production. The latter could be the case for VOT in Austrian German. The tendency for deaspiration may lead to fuzzy voicing categories, blurred perceptual boundaries and a reduced ability to distinguish between “voiced” and “voiceless” sounds based on VOT duration. Alternatively, it could also be that the tendency for deaspiration results in a perceptual boundary shift, and that Austrian listeners perceive stops as voiceless at shorter VOT durations than German listeners.

## 2. METHODS

In order to investigate the perceptual relevance of VOT for Austrian German listeners and their phonemic categorisation abilities, a listening experiment was conducted where subjects had to identify semi-manipulated speech stimuli varying in their VOT duration.

### 2.1. Subjects

In total, 33 native adult Austrian listeners (age span: 21-65) from Styria (located in the southeast of Austria) took part in the perception study. Additionally, 47 native German listeners (age span: 19-60) served as a control group. The German listeners all lived in Styria but were born and raised in various regions of Germany. They were subdivided into “South German” (n = 30) and “North German” listeners (n = 17) based on their place of origin.

All participants were raised monolingual, reported normal hearing and were naïve to the purpose of the study.

Testing was conducted in one session lasting approximately 45 minutes and subjects received 20 Euro for their participation.

## 2.2. Stimuli and experimental procedure

The experimental material consisted of natural speech tokens (the real German words “backen” [*to bake*], “packen” [*to pack*], “danken” [*to thank*], “tanken” [*to fuel*], “Gasse” [*alley*], “Kasse” [*cash point*]) produced by a female native Austrian speaker which were acoustically manipulated regarding their VOT duration. The stimuli formed six continua (one for each word) and varied from 0-60ms VOT in 5ms steps (yielding 13 different stimuli for each word).

While all six continua had an identical VOT range, the continua varied in their burst identity and vowel onset characteristics (i.e. ‘lenis’ bursts and vowel onsets in those stimuli manipulated from original “backen”, “danken” and “Gasse” and ‘fortis’ bursts and vowel onsets in those stimuli manipulated from “packen”, “tanken” and “Kasse”). This design was chosen to test if and how burst information and vowel onset characteristics affect VOT boundaries in German-speaking listeners.

Except for VOT manipulation, acoustic interference was kept at a minimum and no spectral or temporal changes were conducted besides a levelling of burst durations at 5ms.

Stimuli were auditorily presented over headphones and subjects had to perform a two-alternative forced choice (2AFC) identification task. In total, subject were ask to identify the voicing category of 468 stimuli (6 continua \* 13 VOT steps \* 6 repetitions).

All acoustic manipulations as well as stimulus presentation were performed in *Praat* [3].

## 3. ANALYSIS

For each subject, so-called psychometric curves were constructed which model the probability of “voiceless”-responses as a function of VOT duration. This was done by means of *logistic regression analysis* (see formula below), conducted separately for each continuum and each subject. The fitting of the models was performed in *R* using the ‘quickpsy’ package [9].

$$(1) \quad P(Y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

The location of the perceptual boundary between the two voicing categories was defined as the 50%-crossover point in the continuum, i.e. the VOT duration for which the predicated probability of a “voiceless”-response and a “voiced”-response was

equal. If a subject did not reach this crossover point within the presented VOT range, their data were excluded from analysis.

In addition to the location of the VOT boundaries, the “steepness” of the curves, measured by means of the slope parameter  $\beta_1$  of the logistic regression models, was also taken into consideration.

In speech perception research, the steepness of the psychometric curve is generally interpreted as a measure for the preciseness of categorical perception (see e.g. [6]). The steeper the curve, the more abruptly listeners shift from “voiced” to “voiceless” responses, indicating a sharp perceptual boundary. When the curve is less steep, however, this suggests a less categorical (and more continuous or even random) perception of the presented stimuli.

The predicted boundary location as well as the slope values of each subject were used as dependent variables in one-way ANOVAs with the between-subjects factor “listener group”. Analysis was done for each of the six continua separately.

## 4. RESULTS

### 4.1. VOT boundary location

Overall, the location of the VOT boundary was very similar across listener groups (see Table 1).

**Table 1:** Mean boundary location (in ms) and standard deviation for each continuum and listener group.

	Austrian	South German	North German
“backen”	32.58 (SD=6.69)	29.85 (SD=4.81)	30.10 (SD=4.40)
“danken”	28.55 (SD=9.13)	31.51 (SD=8.57)	33.38 (SD=12.20)
“Gasse”	42.95 (SD=5.16)	42.57 (SD=7.71)	41.05 (SD=7.50)
“packen”	16.93 (SD=10.22)	12.90 (SD=5.67)	12.79 (SD=6.42)
“tanken”	12.81 (SD=7.35)	13.53 (SD=6.59)	14.43 (SD=7.27)
“Kasse”	19.91 (SD=6.34)	15.70 (SD=4.52)	15.72 (SD=7.67)

Statistical analysis did not show significant shifts in boundary locations between the three listener groups. The only exception was “Kasse”. Here, Austrian listeners needed on average a VOT duration that was 4ms longer to give a “voiceless”-response than German listeners ( $F(2, 75) = 4.537, p = 0.0138$ ).

The listener effect for “Kasse” and especially the finding that Austrian listeners needed a *longer* VOT than German listeners may be surprising at first.

Word-initial prevocalic velars are usually not des aspirated in neither the standard variety of German spoken in Germany nor in Austria [8]. Therefore, we would actually not expect a listener effect for this continuum. However, in several production studies on Austrian German, a tendency for affrication of the voiceless velar plosive has been attested [11]. This could be the reason for Austrian listeners needing a longer VOT to perceive a velar plosive as “voiceless”, since they might be used to a slightly longer VOT due to affrication.

The results for the other five continua suggest that Austrian listeners need on average the same VOT duration to perceive a stimulus as “voiceless” as German listeners.

#### 4.2. Steepness of the curve

In order to give an estimate of the steepness of the perceptual boundary, the slope parameter  $\beta I$  was calculated for each subject and continuum. Mean values are given below (see Table 2).

**Table 2:** Mean slope value and standard deviation for each continuum and listener group.

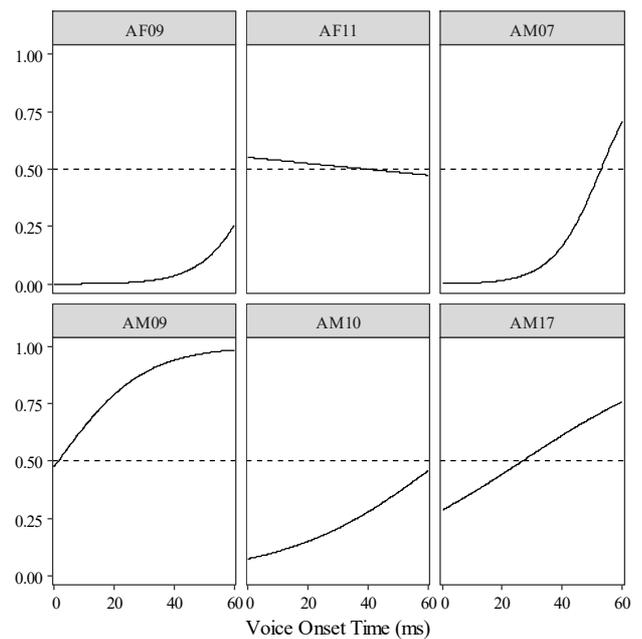
	Austrian	South German	North German
“backen”	0.28 (SD = 0.14)	0.62 (SD = 0.98)	0.35 (SD = 0.14)
“danken”	0.20 (SD = 0.11)	0.30 (SD = 0.55)	0.22 (SD = 0.08)
“Gasse”	0.54 (SD = 0.74)	0.50 (SD = 0.74)	0.55 (SD = 0.73)
“packen”	0.42 (SD = 0.76)	0.25 (SD = 0.10)	0.65 (SD = 1.02)
“tanken”	0.47 (SD = 0.78)	0.28 (SD = 0.15)	0.28 (SD = 0.12)
“Kasse”	0.40 (SD = 0.65)	0.28 (SD = 0.14)	0.49 (SD = 0.78)

In neither case, the differences between listener groups reached statistical significance. This might indicate that the perceptual boundaries were as sharp in Austrian as in German listeners. Based on these results, there is no empirical evidence to claim a reduced ability for voicing categorisation in Austrian listeners.

While this might be true at a group level, it has to be mentioned that there was a considerable amount of individual differences between subjects. As mentioned before, some subjects had to be excluded since they did not reach a 50%-crossover point. Additionally, there were several subjects who did not reach 25%- or 75%-crossover points. This means that within the tested VOT range, the probability of a

“voiceless”-response never exceeded 0.75 – or, in some cases, fell below 0.25. In other words, the voicing category of the presented stimuli was never perceived unambiguously. As a result, the psychometric curves of some subjects deviated from the classic sigmoid shape that is characteristic for categorical perception. Interestingly, for “backen” and for “danken” only Austrian subjects showed atypical identification curves (see Figure 1 for the “danken”-curves of these subjects).

**Figure 1:** Non-sigmoidal psychometric curves for “danken” stimuli.



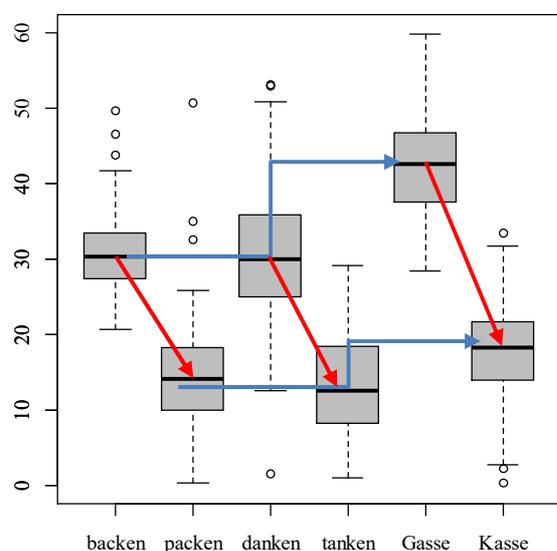
2x2 Fisher’s exact tests showed that the number of subjects with atypical identification curves differed significantly between listener groups. For “backen” and “danken” statistically significant more Austrian subjects than German subjects showed non-categorical responses (“backen”:  $p = 0.02587$ ; “danken”:  $p = 0.003686$ ). This might indicate a reduced ability for voicing categorisation based on VOT duration in these individual listeners.

#### 4.3. Effect of place of articulation

According to the literature, place of articulation can shift the location of the perceptual VOT boundary. Boundary location is said to be shortest for bilabial continua, longest for velar continua and intermediate for alveolar continua [1].

A place of articulation effect on boundary location could also be observed in our data (see Figure 2). Since Austrian and German participants did not differ in this regard, results are pooled across listener groups.

**Figure 2:** Effect of place of articulation (blue arrows) and burst/vowel onset characteristics (red arrows) on the VOT boundaries (pooled across listener groups).



Velars were perceived as “voiceless” at longer VOT durations than bilabials or alveolars. This was true for lenis stimuli ( $F(2, 148) = 101.3, p < 0.001$ ) as well as for fortis stimuli ( $F(2, 144) = 14.06, p < 0.001$ ). For both, lenis and fortis stimuli, post-hoc t-tests (with Bonferroni correction) revealed no statistically significant difference between the boundary location of bilabials and alveolars but only between bilabials and velars ( $p < 0.001$ ) and between alveolars and velars ( $p < 0.001$ ).

#### 4.4. Effect of burst and vowel onset characteristics

When the boundary locations in the lenis and fortis continua are compared with each other, a strong effect of burst identity and/or vowel onset characteristics becomes evident. Stimuli that were designed from original ‘fortis’ onsets (i.e. “packen”, “tanken” and “Kasse”) needed a strikingly shorter VOT to be perceived as “voiceless” than stimuli designed from ‘lenis’ onsets (i.e. “backen”, “danken”, “Gasse”). At VOT durations well below 20ms (which equals an aspiration duration of only 15ms, since the burst duration was set to 5ms), fortis stimuli were perceived as “voiceless” (see Figure 2).

These results indicate that burst and/or vowel onset characteristics that correlate with the lenis/fortis distinction play a crucial role for voicing perception in German-speaking listeners. A literally unaspirated stimulus is likely to be perceived as “voiceless” if its burst and vowel onset features are those typical for fortis stops.

This probably challenges the idea that VOT is the primary cue for voicing perception in German and Austrian listeners.

However, it is not clear yet, whether burst or vowel onset characteristics – or both – are responsible for this effect. Hence, a perception experiment in which these two features are varied independently from each other would be necessary (e.g. with cross splicing fortis and lenis bursts and fortis and lenis vowel onsets).

## 5. SUMMARY

Overall, the tendency to reduce aspiration in Austrian German does not seem to affect the location and sharpness of the perceptual VOT boundaries in Austrian German listeners. Austrian and German listeners did not show different perception results for most of the presented VOT continua. The only exception was the “Kasse” continuum for which Austrian listeners showed a statistically significant boundary shift and needed a longer VOT duration to label a stimulus as “voiceless” than German listeners. This could probably be explained by the fact that velars tend to be affricated in Austrian German stop production [11].

However, on an individual level, considerable differences regarding the degree of categoricity of perception were found. Several subjects showed psychometric curves that diverged from the classic sigmoidal form. For “backen” and “danken” stimuli, these subjects were exclusively Austrian listeners. Since the overall shape of an identification curve can be an indicator for the categorisation abilities of the listener (with a sigmoidal shape indicating “classic” categorical perception [5]), one could argue that these listeners did not perceive VOT as categorical as the rest of the subjects. Given the small sample size, the claim that this was due to the tendency for deaspiration in Austrian German would be premature. However, it is somewhat striking that for “backen” and “danken”, only Austrian subjects showed atypical identification curves.

Leaving individual differences aside, results also indicate a place of articulation effect on VOT boundary locations, with velars needing the longest VOT to be perceived as “voiceless”.

Additionally, burst and/or vowel onset characteristics were shown to strongly affect the location of the VOT boundaries. This could be interpreted as evidence that VOT is not the only (and probably not the primary) acoustic-phonetic cue for voicing perception in German-speaking listeners.

## 6. REFERENCES

- [1] Abramson, A., Lisker, L. 1970. Discrimination along the voicing continuum: cross-language tests. *Proc. 6<sup>th</sup> ICPHS Prague*, 569–573.
- [2] Beckman, J., Jessen, M., Ringen, C. 2013. Empirical evidence for laryngeal features: aspirating vs. true voice languages. *Journal of Linguistics* 49 (2), 259–284.
- [3] Boersma, P., Weenink, D. 2017. *Praat: doing phonetics by computer*. Version 6.0.28, <http://www.praat.org/>
- [4] Grassegger, H. 1988. *Signalphonetische Untersuchungen zur Differenzierung italienischer Plosive durch österreichische Sprecher* (= Forum Phonetikum 40). Hamburg: Buske.
- [5] Hazan, V., Wilson, G., Howells, D., Miller, D., Abberton, E., Fourcin, A. 1995. Speech pattern audiometry for clinical use. *European Journal of Disorders of Communication* 30, 116–123.
- [6] Heimrath, K., Fischer, A., Heinze, H.-J., Zaehle, T. 2016. Changed categorical perception of consonant-vowel syllables induced by transcranial direct current stimulation (tDCS). *BMC Neuroscience* 17:8.
- [7] Jessen, M. 1998. *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam/Philadelphia: Benjamins.
- [8] Krech, E.-M., Stock, E., Hirschfeld, U., Anders, L. C. (eds). 2009. *Deutsches Aussprachewörterbuch*. Berlin/New York: de Gruyter.
- [9] Linares, D., López-Moliner, J. 2016. quickpsy: an R package to fit psychometric functions for multiple groups. *The R Journal* 8 (1), 122–131.
- [10] Lisker, L., Abramson, A. S. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20, 384–422.
- [11] Moosmüller, S., Ringen, C. 2004. Voice and aspiration in Austrian German plosives. *Folia Linguistica*, 38, 43–62.
- [12] Muhr, R. 2001. Varietäten des Österreichischen Deutsch. *Revue Belge de Philologie et d'Histoire* 79, 779–803.