Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication

Session 2pSCa: Variability in Speech Intelligibility: Behavioral and Neural Perspectives

2pSCa2. The impact of variation in phoneme category structure on consonant intelligibility

Valerie Hazan*, Rachel Romeo and Michèle Pettinato

***Corresponding author's address: Speech Hearing and Phonetic Sciences, University College London, Chandler House, London, WC1N 1PF, Greater London, United Kingdom, v.hazan@ucl.ac.uk**

Newman et al. [J. Acoustic. Soc. Am, 109, 1181-1196 (2001)] suggested that phoneme identification accuracy and speed for a given talker was affected by the degree of variability in their production of phoneme categories. This study investigates how intra-talker variability in the production of two phoneme contrasts varies with age and gender, and how this variability affects perception. Multiple iterations of tokens differing in initial consonants (/s/-/ʃ/, /p/-/b/) were collected via picture elicitation from 40 adults and 31 children aged 11 to 14; measures of within-category dispersion, between-category distance, overlap and discriminability were obtained. While females produced more discriminable categories than males, children produced farther yet more dispersed - and thus similarly discriminable - categories than adults. Variability was contrast-specific rather than a general talker characteristic. Tokens with initial /s/-/ʃ/ from pairs of adult and child talkers varying in between-category distance or overlap were presented for identification. The presence of overlap had a greater effect on identification accuracy and speed than between-category distance, with strongest effects for adult speakers, but reaction time correlated most highly with within-category dispersion. These data suggest that talkers who are less consistent in their speech production may be perceived less clearly than more internally-consistent talkers.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4800618]
Received 21 Jan 2013; published 2 Jun 2013

INTRODUCTION

Individual talkers may vary greatly in their inherent intelligibility; this can be reflected in substantial variations in intelligibility rates when the same words or sentences produced by a range of talkers are presented for identification (e.g., Bradlow, Torretta and Pisoni, 1996; Hazan & Markham, 2004). Studies have sought to identify which talker characteristics are significantly related to intelligibility (Bradlow et al., 1996; Hazan & Markham, 2004; van Dommelen & Hazan, 2012), the implicit assumption being that the greater the acoustic-phonetic distance between phoneme categories, the easier these categories should be discriminated by listeners. However, Newman, Clouse & Burnham (2001) proposed that perceptual processing may be affected not so much by between-category distance as by the degree of variability within a phoneme category and presence/absence of overlap between contrasting categories. For a small set of talkers, they evaluated intra-talker variability by measuring spectral distance for multiple iterations of the voiceless sibilant fricatives /s/ and /ʃ/; they also analyzed the degree of overlap between category distributions. In identification experiments using tokens selected from talkers varying in their degree of distance and overlap, they found that listeners were slower to identify initial consonants in tokens spoken by more internally-variable talkers, with category overlap affecting perception above and beyond between-category distances. Hazan & Baker (2011a) examined similar measures for /p/-/b/ and /s/-/ʃ/contrasts and also found a wide range of variability in between-category distance and within-category dispersion amongst talkers; however, in perception tests, they did not find that the more intelligible talkers were the least variable ones. Methodological differences between the perception tests carried out in these two studies (e.g. difference in the number of talkers, presence of background noise) could have led to these conflicting findings.

This study extended these two studies, and also reduced methodological differences between them. As in Hazan and Baker (2011a), two contrasts were included to investigate the consistency of intra-talker variability across phoneme contrasts, and the study extended the investigation of intra-talker variability to child speech by including participants aged 11-14 years. The previous literature suggests that children typically show greater variability in speech production until teenage years (Lee, Potamianos and Narayanan, 1999). Finally, this study systematically analyzed gender differences. Our second goal was to investigate how and to what extent intra-talker variability influences listeners' accuracy and speed of processing when identifying initial /s/-/ʃ/consonants. Pairs of adult and child talkers with extreme values of category overlap and between-category distance were included in the identification task to investigate whether these acoustic-phonetic properties internal to phonemic category structure affect speech perception.

EXPERIMENT 1: PRODUCTION

Speech was recorded from native Southern British English talkers with no reported history of speech or hearing impairment. There were 40 adults (20 M, aged 18-29 yrs, mean 20.8; 20 F, aged 20-28 yrs, mean 23.5) and thirty-one children aged 11-14 years (14 M, mean 13.1; 17 F, mean 13.6). Words were elicited via a picture-naming task: as each picture appeared on a computer screen, participants named it aloud within the frame 'I can see a (noun)'. Eight near-minimal pairs were included in the analyses: four containing word-onset phonemes /s/ or /ʃ/ (sea-sheep, seat-sheet, cell-shell and sack-shack) and four containing /p/ or /b/ (peach-beach, pea-bee, pill-bill, pin-bin). Each picture was presented 8 times in a randomized order yielding, for each talker, 32 tokens per phoneme category.

Consonant onsets and offsets were annotated for each token, and Praat (Boersma & Weenink, 2012) scripts were used to calculate fricative duration and center of gravity (or centroid, i.e. mean frequency weighted by amplitude) for /s/-/ʃ/ tokens, and Voice Onset Time (VOT) for /p/-/b/ tokens. Four additional measures were derived for each phoneme contrast per talker: within-category dispersion (the mean of the standard deviations of centroids for fricative tokens and VOT for stop tokens), between-category distance (difference between the means for each category), category distance/overlap (difference between the minimum /s/ and maximum /ʃ/ centroids, and between the minimum /p/ and maximum /b/ VOTs), and an overall measure of phoneme discriminability d(a) (difference between the mean centroids or VOT divided by the square root of the mean of the variances).

Gender and age effects on centroid values (for the fricative contrast) and VOT values (for the stop contrast) were broadly as expected from the literature. Higher centroid values were obtained for female than male talkers, consistent with Jongman et al. (2000); for the stop voicing contrast, gender effects just failed to reach significance; this is consistent with findings of Ryalls et al. (1997) and Morris et al., (2008) although a gender effect had been

found by Whiteside & Irving (1998). Children aged 11-14 produced /s/ with higher centroids than adults, but values for /ʃ/ did not differ across groups.

TABLE 1. Mean values for between-category distance, within-category dispersion and discriminability for the /s/-/ʃ/ and /p/-/b/contrasts. These values are presented for the four talker groups varying in age and gender. Standard deviations are in parentheses.

	/s/-/ʃ/			/p/-/b/		
	Between-cat distance (Hz)	Within-cat dispersion (Hz)	Discrim. d(a)	Between-cat distance (ms)	Within-cat. dispersion (ms)	Discrim. d(a)
Adult Male (N=20)	2141 (456)	361 (55)	5.97 (1.6)	45 (13)	8 (3)	4.68 (1.09)
Adult Female (N=20)	2952 (662)	348 (70)	8.66 (2.4)	57 (13)	9 (3)	5.61 (1.54)
Child Male (N=14)	2326 (776)	476 (122)	4.92 (1.8)	65 (22)	14 (4)	4.02 (1.18)
Child Female (N=17)	3688 (824)	457 (73)	8.02 (2.3)	70 (19)	12 (3)	5.05 (1.64)

TABLE 2. F values, significance and observed power values resulting from analyses of variance on the between-category distance, within-category dispersion and discriminability measures, with gender and age as between-subject factors.

	/s/-/ʃ/	/p/-/b/
	Between-category distance	
gender	$F(1,67)=44.3$, $p<0.001$, $\eta^2=0.398$	$F(1,67)=4.76$, $p<0.05$, $\eta^2=0.066$
age	$F(1,67)=8.0$, $p<0.01$, $\eta^2=0.106$	$F(1,67)=16.8$, $p<0.001$, $\eta^2=0.200$
	Within-category dispersion	
gender	$F(1,67)=0.73$, N.S.	$F(1,67)=0.65$, N.S.
age	$F(1,67)=34.0$, $p<0.001$, $\eta^2=0.337$	$F(1,67)=29.6$, $p<0.001$, $\eta^2=0.306$
	Discriminability d(a)	
gender	$F(1,67)=34.2$, $p<0.001$, $\eta^2=0.338$	$F(1,67)=8.7$, $p<0.005$, $\eta^2=0.115$
age	$F(1,67)=2.9$, N.S.	$F(1,67)=3.4$, N.S.

The focus of the study was on age and gender effects on between-category distance, within-category dispersion and contrast discriminability (See Table 1 for mean values and Table 2 for a summary of statistical analyses). Females showed greater between-category distance than males, and children than adults, mostly due to higher /s/ centroids and longer /p/ VOTs. In terms of within-category dispersion, males and females did not differ for either contrast, yet children exhibited more dispersed categories than adults, indicating a higher level of within-category acoustic variability for children. While this spectral fricative measure has yet to be analyzed for adolescents, this is consistent with findings of Munson (2004) for younger children and with Lee et al. (1999) in terms of temporal fricative measures for children up to age 14, suggesting children's dispersion in productions of /s/ and /ʃ/ does not decrease to adult-like values until after age 14.

In terms of overall discriminability, males' categories were less discriminable than females' for both fricatives and stops, and for both adults and children. This is consistent with higher intelligibility rates typically obtained for female talkers in word and sentence intelligibility tests (Bradlow et al., 1996; Hazan & Markham, 2004). However, children's greater between-category distance was sufficient to counteract larger dispersions to yield only marginally lower overall category discriminability d(a) than adults.

In terms of overlap between category distributions (values not shown), males in both the child and adult groups were exclusively responsible for overlap between fricative categories (2 adults, 5 children), while some talkers of both genders exhibited overlap between stop categories (1 man, 3 boys, 2 women, 1 girl). Overlap was more pronounced for fricatives (up to 50% of a talker's tokens) than for stops (maximum 14% of tokens).

EXPERIMENT 2: PERCEPTION

The aim of the perception experiment was to evaluate the effect of between-category distance and category overlap on the accuracy and speed of initial consonant identification. The listener group consisted of 22 right-handed adult native talkers of British English with normal hearing (10 M, 12 F, ages: 19-51 yrs, mean: 26). Pairs of talkers were chosen from both age groups and representing variability in either overlap or distance, yielding four conditions: adult overlap, child overlap, adult distance, and child distance. For each condition, one talker exhibited very close but non-overlapping categories, while the other exhibited either a high magnitude of phoneme overlap or a high magnitude of distance between phonemes. Within-category dispersion was kept as similar as possible across talkers. Each pair of talkers was of the same gender: male for overlap conditions and female for distance conditions. All 64 /s/-/ʃ/ tokens (32 per phoneme) from each talker were presented.

Participants were tested individually in a sound-treated booth. Tokens were presented binaurally through noise-canceling headphones. Each condition (adult overlap, adult distance, child overlap, child distance) constituted a block, within which all tokens from both talkers were randomized. Block presentation was counterbalanced across listeners. Listeners were instructed to identify the initial consonant of the word as either an "s" or "sh" as quickly and accurately as possible, even before the word ended, by pressing the corresponding key marked on the computer keyboard. Response times were measured from stimulus onset. Mean accuracy rates and response times for correct trials were calculated per participant per block, after trials with RTs greater than 2 standard deviations outside participant means were excluded.

Similar to findings of Newman et al. (2001), listeners were highly accurate at discriminating phonemes, with only slight differences between talkers, indicating listeners were ultimately able to overcome any ambiguity a talker's variability might pose to intelligibility. However, the costs of parsing an utterance by the more variable talkers are evident in response times (see Figure 1). There was no main effect of phoneme [$F(1,21)=2.133$, $p=0.159$] across all conditions: neither fricative was more or less difficult to categorize than the other, even though acoustic measurements had revealed generally greater within-category dispersion for /s/ than for /ʃ/. A main effect of age [$F(1,21)=6.306$, $p<0.05$, $\eta^2=0.231$] indicated longer response times to child talkers than adults. A follow-up ANOVA indicates this is driven by differences in overlap conditions [$F(1,21)=7.328$, $p<0.05$, $\eta^2=0.259$]. Also, a main effect of variability type [$F(1,21)=5.105$, $p<0.05$, $\eta^2=0.196$] indicated longer response times in overlap conditions than distance conditions. A follow-up ANOVA indicates this is driven by differences in child conditions [$F(1,21)=5.116$, $p<0.05$, $\eta^2=0.196$], with longer RTs in the overlap conditions, and no effect of variability type in adult conditions [$F(1,21)=0.267$, $p>0.6$]. Therefore, talkers whose /s/ and /ʃ/ categories overlapped in frication centroids prompted slower responses, suggesting listeners do indeed use frication centroids as a cue to phonetic categorization, and when non-discrete categories render this cue insufficient, listeners must rely on other cues, resulting in slower parsing. This replicates the results of Newman et al. for adults and extends this finding developmentally. It is notable that the effect size was smaller for child talkers, even though the children exhibited a higher degree of overlap and dispersion than the adult talker pairs. Since listeners had accurately reported discriminating child talkers from adults, this perhaps suggests listeners differentially approach variability for talkers of different age groups. In other words, when listeners encounter a child's voice, their perceptual mechanisms might expect children to have more natural variability and thus automatically take account of alternative cues, which in itself may slow overall processing of child voices.

In addition to the presence of category overlap, the magnitude of distance between non-overlapping categories also affected perception. Listeners responded faster to adult talkers with larger distances between categories, suggesting greater category differentiation facilitates phonemic parsing. However, this effect size was smaller than in the adult overlap condition, suggesting the mere presence of overlap in a talker's categories affects the speed of perception over and above the magnitude of distance between them. For child talkers differing in distance, responses to the /s/ tokens spoken by the supposedly more intelligible high-distance talker were slower, in contrast to the pattern for adult talkers. However, accuracy rates were also higher for the high distance talker and there was therefore a speed-accuracy trade-off, in which listeners are marginally more accurate at identifying tokens by the high distance child talker, but do so significantly more slowly. While this renders it difficult to interpret talker effects for this condition, it is possible that the trade-off masks the same pattern seen for age groups in overlap conditions.

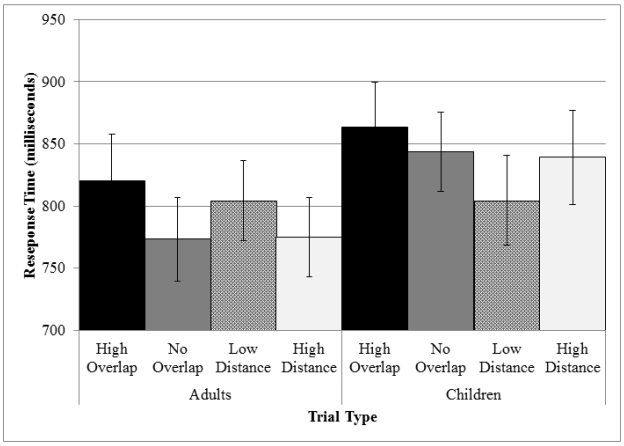

FIGURE 1. Mean response times (in milliseconds) for the 'overlap' and 'distance' conditions of the /s/-/ʃ/ identification task for the adult-talker and child-talker pairs, averaged over /s/ and /ʃ/ responses. Error bars represent standard error.

Talkers selected for the study did not vary widely in category dispersion but it was not possible to totally control this factor across talkers, so the effect of variations in dispersion were also examined by two-tailed Pearson's product-moment correlations. Neither distance, magnitude of overlap or discriminability correlated with accuracy or response times (all $|r|<0.646$, all $p>0.08$). However, talker dispersion was correlated with listeners' RTs, both overall ($r=0.782$ $p<0.05$) and with the one outlying child talker M16 removed ($r=0.814$, $p<0.05$), such that talkers with larger within-category dispersion prompted slower listener RTs. This correlation remained even when controlling for between-category distance ($r=0.801$, $p<0.05$), and even marginally remained when controlling for overall discriminability ($r=0.744$, $p<0.06$), suggesting talkers' intelligibility is strongly linked to their within-category dispersion, over and above the distance between those phonemes.

CONCLUSION

The production analysis showed that within-category dispersion is still greater for children aged 11-14 yrs than for adults, suggesting that some development is still ongoing within that age range. However, the greater between-category distance seen for children counteracted the effect of this greater within-category dispersion, as shown by the lack of a significant age effect for discriminability measures.

The finding that a talker's within-category dispersion was a significant factor in the ease of processing of their speech contrasts has implications for our understanding of the factors that influence talker intelligibility. Studies that have sought acoustic-phonetic correlates of talker intelligibility when talkers were producing conversational speech (Bradlow et al., 1996; Hazan and Markham, 2004) or a clear speaking style (e.g., Hazan and Baker, 2011b) have typically correlated intelligibility scores with mean values for various acoustic-phonetic measures, while taking no account of the degree of within-category dispersion for individual talkers. This may be one reason why correlations between specific acoustic-phonetic measures and intelligibility rates are typically found to be weak, although another reason for this lack of strong correlation is the fact that individual talkers may be using different strategies when enhancing their speech (Hazan and Markham, 2004). Focusing greater attention on measures of within-category dispersion may help to clarify the relation between acoustic-phonetic characteristics of an individual talker's productions and their intelligibility.

ACKNOWLEDGMENTS

This work was supported by the Economic and Social Research Council [grant number: RES-062-23-3106].

REFERENCES

Boersma, P. & Weenink, D. (**2012**). "Praat: doing phonetics by computer" [Computer program]. Version 5.3.07, http://www.praat.org/. (last retrieved 5 March 2012.)

Bradlow A. R., Torretta G. M., & Pisoni D. B. (**1996**). "Intelligibility of normal speech. 1. Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Hazan, V. & Baker, R. (**2011a**). "Is Consonant Perception Linked to Within-Category Dispersion or Across-Category Distance?," Proc. 17th Int. Cong. Phonetic Sc., 839-842.

Hazan, V. and Baker, R. (**2011b**). "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," J. Acoust. Soc. Am. **130**, 2139-2152.

Hazan, V. & Markham D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Acoust. Soc. Am. **116**, 3108–3118.

Jongman, A., Wayland, R., & Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252-1263.

Lee, S., Potamianos, A., & Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455-1468.

Morris, R. J., McCrea, C. R., & Herring, K. D. (**2008**). "Voice onset time differences between adult males and females: Isolated syllables," J. Phon. **36**, 308-317.

Munson, B. (**2004**). "Variability in /s/ production in children and adults: evidence from dynamic measures of spectral mean," J. Speech Lang. Hear. Res. **47**, 58-69.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (**2001**). "The perceptual consequences of within-talker variability in fricative production," J. Acoust. Soc. Am. **109**, 1181-1196.

Ryalls, J., Zipprer,A., Baldauff, P. (**1997**). "A preliminary investigation of the effects of gender and race on Voice Onset Time," J. Speech Hear. Res. **40**, 642-645.

Van Dommelen, W., & Hazan, V. (**2012**). "Impact of talker variability on word recognition in non-native listeners," J. Acoust. Soc. Am. 132, 1690-1699.

Whiteside, S. P. & Irving, C. (**1998**). "Speakers' sex differences in voice onset time: A study of isolated word production" Percept. Motor Skill. **86**, 651–654.