HAL open science

Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners

Monja A. Knoll, Maria Uther, Alan Costall

▶ To cite this version:

Monja A. Knoll, Maria Uther, Alan Costall. Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners. Speech Communication, 2009, 51 (3), pp.210. 10.1016/j.specom.2008.08.001 . hal-00499232

HAL Id: hal-00499232
https://hal.science/hal-00499232

Submitted on 9 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners

Monja A. Knoll, Maria Uther, Alan Costall

PII: S0167-6393(08)00128-3
DOI: 10.1016/j.specom.2008.08.001
Reference: SPECOM 1744

To appear in: *Speech Communication*

Received Date: 25 October 2007
Revised Date: 6 August 2008
Accepted Date: 7 August 2008

Please cite this article as: Knoll, M.A., Uther, M., Costall, A., Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners, *Speech Communication* (2008), doi: 10.1016/j.specom.2008.08.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners.

[a*]Monja A. Knoll, [b]Maria Uther, & [a]Alan Costall

[a]Department of Psychology, University of Portsmouth, King Henry Building, King Henry I Street, Portsmouth, PO1 2DY, U.K. Email: Monja.Knoll@port.ac.uk; Alan.Costall@port.ac.uk

[b]Centre for Cognition and NeuroImaging, School of Social Sciences, Brunel University, West London, Uxbridge, Middlesex, UB8 3PH, U.K. E-mail: Maria.Uther@brunel.ac.uk

*Corresponding author. Department of Psychology, University of Portsmouth, King Henry Building, King Henry 1 Street, Portsmouth, PO1 2DY. E-mail: Monja.Knoll@port.ac.uk. Fax: +44 (0) 23 9284 6300

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ABSTRACT

Low-pass filtering has been used in emotional research to remove the semantic content from speech on the assumption that the relevant acoustic cues for vocal affect remain intact. This method has also been adapted by recent investigations into the function of infant-directed speech (IDS). Similar to other emotion-related studies that have utilised various levels of low-pass filtering, these IDS investigations have used different frequency cut-offs. However, the effects of applying these different low-pass filters to speech samples on perceptual ratings of vocal affect are not well understood. Samples of natural IDS, foreigner- (FDS) and British adult-directed (ADS) speech were low-pass filtered at four different cut-offs (1200 Hz, 1000 Hz, 700 Hz, 400 Hz), and affective ratings of these were compared to those of the original samples. The samples were also analyzed for mean fundamental frequency (F_0) and F_0 range. Whilst IDS received consistently higher affective ratings for all filters, the results of the adult conditions were more complex. ADS received significantly higher ratings of positive vocal affect than FDS with the lower cut-offs (1000 to 400 Hz), whereas no significant difference between the adult conditions was found in the original and 1200 Hz conditions. No difference between the adult conditions was found for encouragement of attention. These findings show that low-pass filtering leaves sufficient vocal affect for detection by raters between IDS and the adult conditions, but that residual semantic information in filters above 1000 Hz may have a confounding affect on raters' perception.

Keywords

Low-pass filtering; infant-directed speech; foreign-directed speech; vocal affect

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1. Introduction

The voice carries a considerable amount of information about a person's emotional state that is communicated via both semantic and vocal acoustic channels (e.g. Kramer, 1963; Starkweather, 1967). Investigation of the emotional acoustic characteristics of speech requires separation of the semantic channel from the vocal acoustic aspects of speech, and this goal has been a recurring concern in emotional speech research. A common method to separate these components has been to standardize the content of the speech by asking the speakers to repeat written text or a list of numbers (e.g. Davitz and Davitz, 1961). However, the difficulty of obtaining true emotional content with these artificial situations is well understood (e.g. Kramer, 1963). In an attempt to avoid these problems, alternative means of deriving content-free speech such as playing speech samples backwards (e.g. Knower, 1941), foreign speech (e.g. Kramer, 1964) and randomized content-splicing (e.g. Scherer, 1971) have been employed, allowing the potential use of naturally occurring speech. These approaches nonetheless presented a variety of new problems, for example playing speech samples backwards causes reversal of the acoustic intonation contours.

An alternative widely used method of separating the two channels, low-pass filtering, has been adapted from research into speech intelligibility (e.g. French and Steinberg, 1947; Pollack, 1948) on the assumption that such filtering would leave a large number of the emotional acoustic cues of the voice intact (e.g. Soskin and Kaufman, 1961; Scherer et al., 1972). It has been suggested that these cues (e.g. intonation contour, rate of speech, pause and rhythm) should remain virtually unaffected by the removal of the upper frequencies (e.g. Rogers et al., 1971; Scherer, 1971, Scherer et al., 1972), since it is mainly the lower frequencies and tonal quality

ACCEPTED MANUSCRIPT

of the voice that are supposed to be important in communicating the speakers' emotional state (Starkweather, 1967).

Because a key advantage of low-pass filtering is that it can be applied to natural speech samples, one area where this approach has recently been used to investigate affect is research into speech directed to infants (IDS). It is believed that IDS serves distinct linguistic, emotional affective and attentional roles (e.g. Fernald and Simon, 1984; Uther et al., 2003), and a growing body of research has attempted to separate these functions (e.g. Burnham et al., 2002; Fernald and Simon, 1984; Kitamura and Burnham, 2003; Trainor et al., 2000). Methods to accomplish this have included investigating acoustic and perceptual aspects of IDS by comparing IDS to other linguistic (Papousek and Hwang, 1991; Uther et al., 2007) and emotional comparison groups (e.g. Burnham et al., 2002; Trainor et al., 2000). In the case of adult linguistic and emotional comparison groups for IDS, content-free speech is clearly crucial for comparing the semantically less challenging IDS to the semantically complex adult speech. Consequently, low-pass filtering has played a key role in the investigation of the affective content in such studies (e.g. Burnham et al., 2002; Kitamura and Burnham, 2003). Two of these comparative studies (e.g. Burnham et al., 2002; Uther et al., 2007) into IDS, and other investigations into children's vocal emotional understanding (e.g. Morton & Trehub, 2001), have utilized different filters.

The applicability of low-pass filtering for speech research has mainly been assessed by comparing a single filter cut-off with the original speech samples (e.g. Starkweather, 1956), content-spliced speech (e.g. Scherer et al., 1972; van Bezooijen & Boves, 1986), reiterant speech (Friend and Farrer, 1993) and foreign speech (e.g. Kramer, 1964). These and other studies in the area (e.g. Cohen and Starkweather,

ACCEPTED MANUSCRIPT

1961; Milmoe et al., 1967) have also used a wide variety of different filter cut-offs. Although, higher cut-offs have also been used (e.g. 1040 Hz, Caporael, 1981), most of these studies have chosen filter cut-offs below 650 Hz. However, it is not clear how much of the relevant nonverbal information may be lost with the removal of the upper frequencies (Kramer, 1963), and no consensus has emerged as to the optimal filter for removing the semantic content, while at the same time conserving most of the acoustic emotional content.

To our knowledge, there is only one existing study (Ross et al., 1973) that has investigated the effect of different filters (between 600 Hz and 150 Hz) on the recognition of basic emotions. The authors found that the higher filters resulted in good recognition rates, and that only the lowest filter (150 Hz) resulted in a considerable deterioration in the raters' ability to correctly identify the emotions. This study, however, was restricted to a low range of filter cut-offs and used speech samples of actors instructed to play out specific emotions whilst reciting a fixed text, rather than spontaneous emotional speech.

The extent to which a wide and higher range of different levels of low-pass filtering influences perception of affect in natural speech samples therefore remains unknown. Investigation of this factor is crucial, not only for justification of using low-pass filtering in future emotional speech research, but also for validation of previous findings from similar studies. This is clearly pertinent in the area of IDS, where the use of natural speech samples is of great importance. However, this issue is equally important to other areas of speech research including hearing impairment, hearing aid design (e.g. Hogan & Turner, 1998; Baer, *et al*., 2002), speech recognition (e.g. Dubno *et al*., 2005) and the design of automated speech recognition systems (White,

ACCEPTED MANUSCRIPT

1975), where frequency filtering (low-pass, high-pass and band-pass) continues to be used.

Here, we attempt to address these questions by investigating how a series of progressively lower filters influence raters' perceptions of vocal affect. We chose IDS, foreigner- (FDS) and British adult-directed speech (ADS) as comparative speech recipient groups, given the perceived high emotional content of IDS relative to these other types of speech (e.g. Burnham et al., 2002; Uther et al., 2007). This approach allows direct investigation of the effects of filtering across filters and between speech produced by the same speakers directed to different audiences. Using a filter of 1000 Hz, it was previously found that IDS was rated higher on positive vocal affect and encouragement of attention than ADS and FDS, whereas ADS was rated higher than FDS in positive vocal affect, but not in encouragement of attention (e.g. Knoll and Uther, 2004; Uther et al., 2007). We were particularly interested to explore how the use of a lower filter in the same data set (such as the 400 Hz filter of Burnham et al., 2002) would have influenced these affective ratings. To achieve this aim, we used low-pass filters of 400 Hz (Burnham et al., 2002), 700 Hz (as an approximate average of studies that have used filter cut-offs between 650 and 720 Hz; e.g. Pollack, 1948; Scherer et al., 1972), 1000 Hz (Uther et al., 2007), and added a further filter cut-off of 1200 Hz (intended as a bridge between commonly used filters and unfiltered speech), and compared them to the original speech samples. The purpose of the present study, therefore, is to determine at which level of frequency cut-off, if any, low-pass filtering influences the perception of vocal affect.

2. Method

ACCEPTED MANUSCRIPT

2.1 Design

We applied a mixed factorial design with two factors. The first factor consisted of five different levels of low-pass filtered speech samples (400 Hz, 700 Hz, 1000 Hz, 1200 Hz and original; between subject measures), and the second factor consisted of the three speech recipient groups (IDS, FDS and ADS; within subjects). The dependent variables were rated vocal affect, and acoustic measures (mean F_0 (Hz) and F_0 range (Hz)).

2.2 Raters

A total of 117 people took part in the experiment; 26 males and 91 females with a mean age of 26.3 years (*sd* 9.0). Participants were required to have no known hearing impairment and had to be fluent in English. This was clearly stipulated in the instructions, and participants had to be capable of understanding and following complex written technical instructions in English (78.4% of the participants were English native speakers). Preliminary analyses of language background (English as first or second language) and gender of the listeners were found to be non-significant, and showed that these factors did not influence the results. Calculation of intra-class correlation coefficient (two-way random model) showed that inter-rater reliability was high for each filter condition (reliability coefficient range from $\alpha = .936$ (original) to $\alpha = .892$ (400 Hz)).

2.3. Stimuli selection and filtering

An existing data set (Uther et al., 2007) consisting of recordings of ten southern English mothers (mean age 30.7 years) talking to their infants (mean age 37 weeks), a foreign (Chinese) and British adult confederate (both females in their early twenties)

ACCEPTED MANUSCRIPT

formed the basis of the present dataset. For the interactions, the mothers were provided with three toys to retain consistency of conversation content, but otherwise the interactions were natural and spontaneous. The interactions lasted between 5 and 10 minutes, but following previous research (e.g. Burnham et al., 2002; Kitamura & Burnham, 2003; Uther et al., 2007), only approximately 30 seconds of each mother's voice was extracted. Extraneous sounds (e.g. the voice of the confederate) were edited out. Each sample was chosen from approximately the start of the mother's dialogue, but speech occurring during the first 30 seconds was generally rejected since the beginning of each interaction was mostly accompanied by distractive sounds. The speech samples were then subjected to successive low-pass filtering (Hann window) using Praat 4.3.14 (Boersma and Weenink, 2005), to remove frequencies above 400 Hz, 700 Hz, 1000 Hz and 1200 Hz. For all of these, the standard recommendation for smoothing at 100 Hz was employed.

2.4. Set-up of the web-based questionnaire

The questionnaire was delivered via the Internet, hosted by the University of Portsmouth (for validity of Internet obtained ratings see Knoll et al., 2008). We did not use basic emotional categories (e.g. joy, anger) for the investigations of affect, because, in contrast to previous research (e.g. Ross et al., 2007), we had not instructed our speakers to convey any specific emotion. Instead, we adapted a questionnaire from that of Kitamura and Burnham (2003), which has been used to investigate *affect* in natural IDS in several studies (e.g. Burnham et al., 2002; Uther et al., 2007). The questionnaire consisted of four affective scales: 1) positive and 2) negative vocal affect, 3) encouragement of attention, and 4) comforting and soothing (see Kitamura and Burnham, 2003 for a full description of these scales). Ratings were obtained on a

ACCEPTED MANUSCRIPT

five point Likert-scale (1 (not at all) to 5 (extremely)). We focused attention on what seemed to be the most critical scales to compare with previous research (Knoll and Uther, 2004; Uther et al., 2007), namely the scales of positive vocal affect (as an example of an emotional affective scale) and encouragement of attention (as an example of an emotional and communicative scale). We also concluded that both of these scales might be more informative in the context of adult speech than, for instance, ‘comforting and soothing’, which relates more specifically to investigations of IDS.

2.5. Procedure and background information

Raters were recruited via a snowballing approach, whereby messages for volunteers were sent, amongst others, to Universities in the UK. Raters were sequentially assigned to one of the five filter conditions, whereby the first person clicking on the link would be in the original condition, the second person in the 1200 Hz condition and so forth (pseudo-random assignment). However, not all of the participants completed the study. This resulted in an assignment of 24 raters in the original and 700 Hz conditions, 21 raters in the 1000 Hz and 400 Hz conditions, and 27 raters in the 1200 Hz condition. Each rater was required to listen to 30 speech samples (10 IDS, 10 FDS and 10 ADS). The order of presentation of the speech samples was randomized for each of the raters, and they were supplied with two test trials of each filtered condition for familiarization with their respective sound. Raters were instructed to adjust the loudness level to their own comfort and hearing ability, as later speech samples would only be played once to avoid habituation effects.

ACCEPTED MANUSCRIPT

2.6. Acoustic analyses

Different levels of low-pass filtering might lead to a decrease in the measures of some of the acoustic features (e.g. F_0 range[1]), possibly influencing perceptual ratings within each filter. We therefore acoustically analyzed each of the speech samples for mean F_0 (Hz) and F_0 range (Hz). Analyses were carried out with Praat 4.3.14 (Boersma and Weenink, 2005). The pitch floor was set at 75 Hz with a ceiling of 600 Hz.

All of our analyses concentrated on the whole (unmodified) 30 second speech samples, firstly because acoustic analyses (e.g. F1/F2 ratio) of individual target words of these data had already been presented in earlier work (Knoll and Uther, 2004; Knoll et al., 2007; Uther et al., 2007). Secondly, we were interested in the *overall* pattern of acoustic characteristics for each speech sample, and what the raters would be able to perceive during their rating task. Mean F_0 and F_0 range are important acoustic characteristics of basic emotions (Scherer, 2003), and analyses of these might provide some insight into the cues that the raters may have used for their ratings.

3. Results

The data for the two affective scales (positive vocal affect and encouragement of attention) and the acoustic analyses were subjected to a mixed MANOVA with speech recipient groups (IDS, FDS and ADS) as within subjects factor, and filters (400 Hz, 700 Hz, 1000 Hz, 1200 Hz and original) as between subjects factor. Results were further explored with univariate ANOVAs and planned contrasts. Asterisks indicate Greenhouse Geissser adjustment.

[1] F_0 as measured in this study is an average over the 30 seconds varying auditory signal. Therefore higher frequencies may be attenuated by the lower filters and can therefore result in lower mean F_0 and F_0 range values.

ACCEPTED MANUSCRIPT

3.1 Affective scales

We found a significant two way interaction for speech recipient groups and filters ($F_{(16, 333.6)}$ = 2.790, p < .001, Wilks' Lambda = .681, η = .086), with both positive vocal affect ($F_{(6.2, 172.6*)}$ = 2.279, p = .037, η = .075) and encouragement of attention ($F_{(5.4, 151.4*)}$ = 3.627, p = .003, η = .115) contributing to this interaction effect (Fig. 1). This interaction is due to a slight decreasing trend for both IDS and FDS in the scale of positive vocal affect with no significant interaction between these two conditions, whereas ADS exhibits a weak parabolic curve and interacts significantly with IDS (p = .028) and FDS (p = .009). There is no significant interaction between ADS and FDS (both display a slight parabolic curve) in the scale of encouragement of attention, whereas IDS again exhibits a slight decreasing trend, interacting with both ADS (p = .003) and FDS (p = .01).

Fig. 1 about here

We found a significant effect of filters on rated affect ($F_{(8, 222)}$ = 2.080, p = .039, Wilks' Lambda = .865, η = .069). Analysis of each individual dependent variable showed that the five filter groups differed only in terms of the ratings of positive vocal affect ($F_{(4, 112)}$ = 4.843, p = .014, η = .104), but not for the ratings of encouragement of attention. Participants in the original speech condition (mean = 3.551) perceived the speech samples to contain significantly more positive vocal affect than the participants in the low-pass filtered conditions (1200 Hz mean = 3.317, p = .033; 1000 Hz mean = 3.225, p = .006; 700 Hz mean = 3.204, p = .002; 400 Hz mean = 3.235, p = .007). No significant difference was found between the remaining low-pass filtered conditions.

ACCEPTED MANUSCRIPT

We also found a significant effect for speech recipient groups ($F_{(4, 109)}$ = 150.679, Wilks' Lambda = .153, $p < .001$, $\eta = .847$) with both scales contributing to these differences (positive vocal affect: $F_{(1.5, 172.6*)}$ = 423.861, $p < .001$, $\eta = .791$; and encouragement of attention: $F_{(1.4, 151.4*)}$ = 343.624, $p < .001$, $\eta = .754$). IDS achieved significantly higher ratings of positive vocal affect and encouragement of attention than both ADS (for both scales at $p < 0.001$) and FDS (for both scales at $p < .001$). ADS also achieved higher ratings of positive vocal affect than FDS ($p < .001$), however, no significant difference between the adult conditions was observed for ratings of encouragement of attention (Fig. 1).

To explore whether these differences between the speech recipient groups were consistent within each of the five filters, we ran planned contrasts between the speech groups for each of the filters for each of the scales. As can be seen in Table 1, IDS achieved consistently higher ratings of positive vocal affect and encouragement of attention than both adult groups for each of the five filters. ADS only received higher ratings of positive vocal affect than FDS in the 1000 Hz, 700 Hz and 400 Hz filters, but not in the original or the 1200 Hz condition. No difference between the two adult groups in the ratings of encouragement of attention was found for any of the five filters.

Table 1 about here

3.2 Acoustic analyses

There was no significant interaction between speech recipient groups and filter type for the acoustic data. We found a significant effect of filters on acoustic measurements ($F_{(8, 88)}$ = 6.733, $p < 0.001$, Wilks' Lambda = .385, $\eta = .380$), although

ACCEPTED MANUSCRIPT

the five filter groups differed only in terms of measures of F_0 range ($F_{(4, 45)}$ = 15.492, $p < .001$, $\eta = .579$) but not for measures of mean F_0. The measures of F_0 range in the 400 Hz condition were significantly lower than in the remaining filter conditions (at $p < .001$). No other significant differences between the filters were observed (see Fig. 2).

Fig. 2 about here.

We also found a significant effect of speech groups on acoustic measurements ($F_{(4, 42)}$ = 19.967, $p < .001$, Wilks' Lambda = .345, $\eta = .655$). Univariate ANOVAs revealed that both mean F_0 ($F_{(1.1, 47.9*)}$ = 51.425, $p < .001$, $\eta = .533$) and F_0 range ($F_{(1.4, 60.9*)}$ = 15.764, $p < .001$, $\eta = .259$) contributed to the differences (Fig. 2). IDS measures of mean F_0 and F_0 range were significantly higher than both ADS (for both measures at $p < .001$) and FDS (for both measures at $p < .001$). Although, the ADS measures of mean F_0 were significantly higher than FDS ($p = .006$), there was no significant difference between the two adult conditions for F_0 range (see Fig. 2). These findings are broadly comparable with our earlier investigations of mean F_0 and F_0 contour shape in target words, where IDS was found to comprise higher mean F_0 and more exaggerated F_0 contours than both adult conditions (Knoll et al., 2007; Uther et al., 2007).

4. Discussion

The results of our study indicate that different low-pass filters influence raters' perception of positive vocal affect and, to a lesser degree, encouragement of attention. Consistent with the assumption that removal of the upper frequencies may lead to a loss of some of the emotional cues in speech, we found that ratings for IDS in both

ACCEPTED MANUSCRIPT

scales, and ratings for FDS for positive vocal affect, gradually decreased as a function of the cut-off frequency. This was not the case for FDS in encouragement of attention, or for ADS in both scales. We suggest that loss of intelligibility across the filters may partly be responsible for these results, and we discuss this possibility below.

In line with previous research (e.g. Kitamura and Burnham, 2003), we found that IDS was consistently rated more highly (for both scales) than both adult conditions across the filter conditions. This result is unsurprising, as IDS is a highly affective speech register, and previous studies using two different low-pass filter cut-offs (Burnham et al., 2002; Uther et al., 2007) had already reported increased emotional affect in IDS compared to adult-speech groups. We also found increased mean F_0 and greater F_0 range in IDS, whereas neither ADS nor FDS displayed the same exaggeration or increase in F_0, which is concordant with earlier findings (e.g. Fernald and Simon, 1984; Knoll et al., 2007). These acoustic features, which are less affected by filtering, have been reported to have both an emotional- affective and attentional function (e.g. Fernald and Simon, 1984; Knoll et al., 2007), perhaps explaining the consistently higher ratings of IDS in both scales.

The higher ratings of positive vocal affect for ADS compared with FDS reported in Uther et al. (2007) were replicated only in the lower filters (1000 Hz to 400 Hz). It is possible that sufficient semantic content of the original samples remains in the relatively liberal 1200 Hz filter cut-off to provide the raters not only with acoustic cues, but also with some contextual semantic information. This notion is also potentially supported by our findings that the original (but closely followed by the 1200 Hz) condition received significantly higher ratings of positive vocal affect than the remaining conditions.

ACCEPTED MANUSCRIPT

Uther et al. (2007) proposed that the lower ratings of positive vocal affect found in FDS might be due to speaker frustration. Our findings are consistent with these interpretations, particularly if certain aspects of speaker frustration reside only in vocal acoustic features rather than in semantic content. For instance, raters may have taken semantic content into account in the higher filters in their evaluation of positive affect, but were forced to concentrate on acoustic signals in the more unintelligible filters. It is interesting to note that ADS was not affected by this 'semantic effect', as ratings here did not decrease gradually with the cut-off frequencies (in both scales). This suggests that semantic content and acoustic features of natural ADS may be mutually supportive, whereas the same does not apply to FDS or IDS. However, the acoustic signal responsible for this effect is not clear. Although ADS was characterized by higher F_0 (which may be associated with the basic "positive" emotion of joy/elation; Scherer, 2003) in four of the conditions, this was not the case for the 400 Hz condition, where the mean difference in ratings between the two adult conditions was the greatest. It is therefore possible that another acoustic modification of FDS (e.g. speech rate; Biersack et al, 2005) is responsible for the reduced positive affect in FDS, and we are currently investigating this possibility.

In contrast to positive vocal affect, we did not find a significant difference between the adult conditions for the scale of encouragement of attention. It is possible that the speakers did not attempt to gain the attention of their speech partner in the adult interactions (supported by the overall low ratings in this scale in the adult conditions), and that it was therefore not possible for the raters to detect such a difference. We suspect that encouragement of attention may be well suited to investigations involving infants, where acoustic modifications (e.g. greater F_0 range) are of great importance in attracting the attention of the speech recipient (e.g. Knoll et

ACCEPTED MANUSCRIPT

al., 2007), but the scale may be less sensitive in detecting differences between the two adult groups. This assumption would need further investigation of speech samples where speakers are required to engage the attention of their speech partner, for instance, teachers addressing pupils in a classroom.

Our results nonetheless indicate that affective comparisons may be carried out with confidence using filters of 1000 Hz to 400 Hz, but that the 1200 Hz filter may not remove enough of the semantic content for this specific purpose. These findings have important implications for infant research utilizing natural speech samples. For instance, the use or choice of a filter is not necessarily critical if IDS is to be compared to a less emotionally charged adult group, but a more severe filter cut-off (≤ 1000 Hz) may be needed for discrimination between speech directed to adults where affective content is similar.

As natural interactions are normally accompanied by some degree of background noise, and since the intelligibility of filtered speech varies depending on the proportion of background noise (Pollack, 1948; Rogers et al., 1973), selection of the appropriate filter very much depends on the level of intelligibility remaining after that filter has been applied. Our results suggest that 1000 Hz was sufficiently severe to impair intelligibility in the present study, although this might vary for other situations where speech is recorded under optimal conditions (i.e. without background noise in a laboratory; see French and Steinberg, 1947). Future studies should thus include questions designed to investigate raters' perceptions of speech sample intelligibility. Of course, if emotional affect is compared between different speech groups, consistency of filter cut-off between conditions remains crucial to the experimental set-up.

ACCEPTED MANUSCRIPT

In summary, the purpose of the present study was to investigate whether the affective salience of speech is left unaffected when different levels of low-pass filtering are used to eliminate intelligibility. The results show that perception of affect is influenced by different frequency filter cut-offs, but that this influence is relatively small, with the greatest effects occurring in the adult conditions. Low-pass filtering is thus a useful tool in affective speech research, particularly where different types of adult-directed speech are the focus of the investigation. These findings suggest that different filters (≤ 1000 Hz) can yield similar results, but that it is obviously necessary to use the same filter for the comparison of different speech types.

ACKNOWLEDGEMENTS

This research was supported by a grant from the ESRC to Monja Knoll. We thank Pushpendra Singh (Newcastle University, UK), Paul Marshman (Portsmouth University, UK), James Uther (Sydney University, Australia), Axelle Philippon (Portsmouth University, UK) and Stig Walsh (NHM, UK) for technical support and discussion. Two anonymous reviewers are thanked for greatly improving this manuscript.

REFERENCES

Baer, T., Moore, B. C. J., and Kluk, K. (**2002**). Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies. Journal of the Acoustical Society of America 112, 1133-1144.

ACCEPTED MANUSCRIPT

Biersack, S., Kempe, V., Knapton, L., 2005. Fine-tuning speech registers: A comparison of the prosodic features of child-directed and foreigner-directed speech. In: Interspeech-2005, pp. 2401-2405.

Boersma, P., Weenink, D., 2005. Praat: doing phonetics by computer (Version 4.3.14) [Computer programme]. Retrieved June 2005 from http://www.praat.org/.

Burnham, D., Kitamura, C., Vollmer-Conna, U., 2002. What's new pussycat? On talking to babies and animals. Science 296, 1095.

Caporael, L. R., 1981. The paralanguage of caregiving: baby talk to the institutionalized aged. Journal of Personality and Social Psychology 40, 876-884.

Cohen, A., Starkweather, J. A., 1961. Vocal cues to the identification of language. American Journal of Psychology 74, 90-93.

Davitz, J. R, Davitz, L. J., 1961. Nonverbal vocal communication of feeling. Journal of Communication 81, 81-86.

Dubno, J. R., Horwitz, A. R., Ahlstrom, J. B., 2005. Recognition of filtered words in noise at higher-than-normal levels: decreases in scores with and without increases in masking. Journal of the Acoustical Society of America 118, 923-933.

Fernald, A., Mazzie, C., 1991. Prosody and focus in speech to infants and adults. Developmental Psychology 27, 209-221.

Fernald, A., Simon, T., 1984. Expanded intonation contours in mother's speech to newborns. Developmental Psychology 20, 104-113.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Friend, M, Farrer, M. J., 1994. A comparison of content-masking procedures for obtaining judgments of discrete affective states. Journal of the Acoustical Society of America 96, 1283-1290.

French, N. R., Steinberg, J. C., 1947. Factors governing the intelligibility of speech sounds. Journal of the Acoustical Society of America 19, 90-119.

Hogan, C. A., Turner, C. W., 1998. High-frequency audibility: Benefits for hearing-impaired listeners. Journal of the Acoustical Society of America 104, 432-441.

Kitamura, C., Burnham, D., 2003. Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year. Infancy 4, 85-110.

Knoll, M. A., Uther, M., 2004. 'Motherese' and 'Chin-ese': evidence of acoustic changes in speech directed at infants and foreigners. Journal of the Acoustical Society of America 116, 2522.

Knoll, M. A., Uther, M., Costall, A., 2008. Evaluation of the utility of the Internet for emotional-affective speech perception studies. Paper presented at the BPS Annual Conference 2008 in Dublin, (abstracts volume, pp. 232-233).

Knoll, M. A., Walsh, S. A., MacLeod, N., O'Neill, M., Uther, M., 2007. Good performers know their audience! Identification and characterisation of pitch contours in infant and foreigner-directed speech, in: MacLeod, N. (Ed.), Automated taxon recognition in systematics: Theory, approaches and applications, CRC Press/The Systematics Association, Boca Raton, Florida, pp. 299-310.

Knower, F. H., 1941. Analysis of some experimental variations of simulated vocal expressions of the emotions. Journal of Social Psychology 14, 369-372.

ACCEPTED MANUSCRIPT

Kramer, E., 1963. Judgment of personal characteristics and emotions from nonverbal properties of speech. Psychological Bulletin 60**,** 408-420.

Kramer, E., 1964. Elimination of verbal cues in judgments of emotion from voice. Journal of Abnormal Social Psychology 68, 390-396.

Milmoe, S, Rosenthal, R., Blane, H. T., Chafetz, M. E., Wolf, I., 1967. The doctor's voice: postdictor of successful referral of alcoholic patients. Journal of Abnormal Social Psychology 72, 78-84.

Morton, J. B., Trehub, S. E., 2001. Children's understanding of emotion in speech. Child Development 72, 834-843.

Papousek, M., Hwang, S.-F. C., 1991. Tone and intonation in Mandarin babytalk to presyllabic infants: comparison with registers of adult conversation and foreign language instruction. Applied Psycholinguistics 12, 481–504.

Pollack, I., 1948. Effects of high pass and low pass filtering on the intelligibility of speech in noise. Journal of the Acoustical Society of America 20, 259-266.

Rogers, P. L., Scherer, K. R., Rosenthal, R., 1971. Content filtering human speech: A simple electronic system. Behavior Research Method and Instruments 3, 16-18.

Ross, M., Duffy, R. J., Cooker, H. S., Sargeant, R. L. (1973). Contribution of the lower audible frequencies to the recognition of emotions. American Annals of the Deaf 118, 37-42.

Scherer, K. R., 1971. Randomized splicing: a note on a simple technique for masking speech content. Journal of Experimental Research 5, 155-159.

ACCEPTED MANUSCRIPT

Scherer, K. R., 2003. Vocal communication of emotion: A review of research paradigms. Speech Communication 40, 227-256.

Scherer, K. R., Koivumaki, J., Rosenthal, R., 1972. Minimal cues in the vocal communication of affect: judging emotions from content-masked speech. Journal of Psycholinguistic Research 1, 269-285.

Starkweather, J. A., 1956. The communication-value of content-free speech. American Journal of Psychology 69, 121-123.

Starkweather, J. A., 1967. Vocal behavior as an information channel of speaker status, in: Salzinger, K., Salzinger, S. (Eds), Research in verbal behavior and some neurophysiological implications, Academic Press, New York, pp. 253-262.

Soskin, W. F., Kauffman, P. E., 1961. Judgment of emotion in word-free voice samples. Journal of Communication 11, 73-80.

Trainor, L. J., Austin, C. M., Desjardins, R. N., 2000. Is infant-directed speech prosody a result of the vocal expression of emotion? Psychological Science 11, 188-195.

Uther, M., Knoll, M. A., Burnham, D., 2007. Do you speak E-n-g-l-i-s-h? A comparison of foreigner- and infant-directed speech. Speech Communication 49, 2-7.

van Bezooijen, R., Boves, L., 1986. The effects of low-pass filtering and random splicing on the perception of speech. Journal of Psycholinguistic Research 15, 103-417.

White, G. M. 1975. Linear predictive residual analysis compared to bandpass filtering for automatic speech recognition. Journal of the Acoustical Society of America 58, 106.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table captions

Table 1. Mean difference, F-value and significance level between IDS, ADS and FDS for each of the five filter conditions for positive vocal affect and encouragement of attention.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1

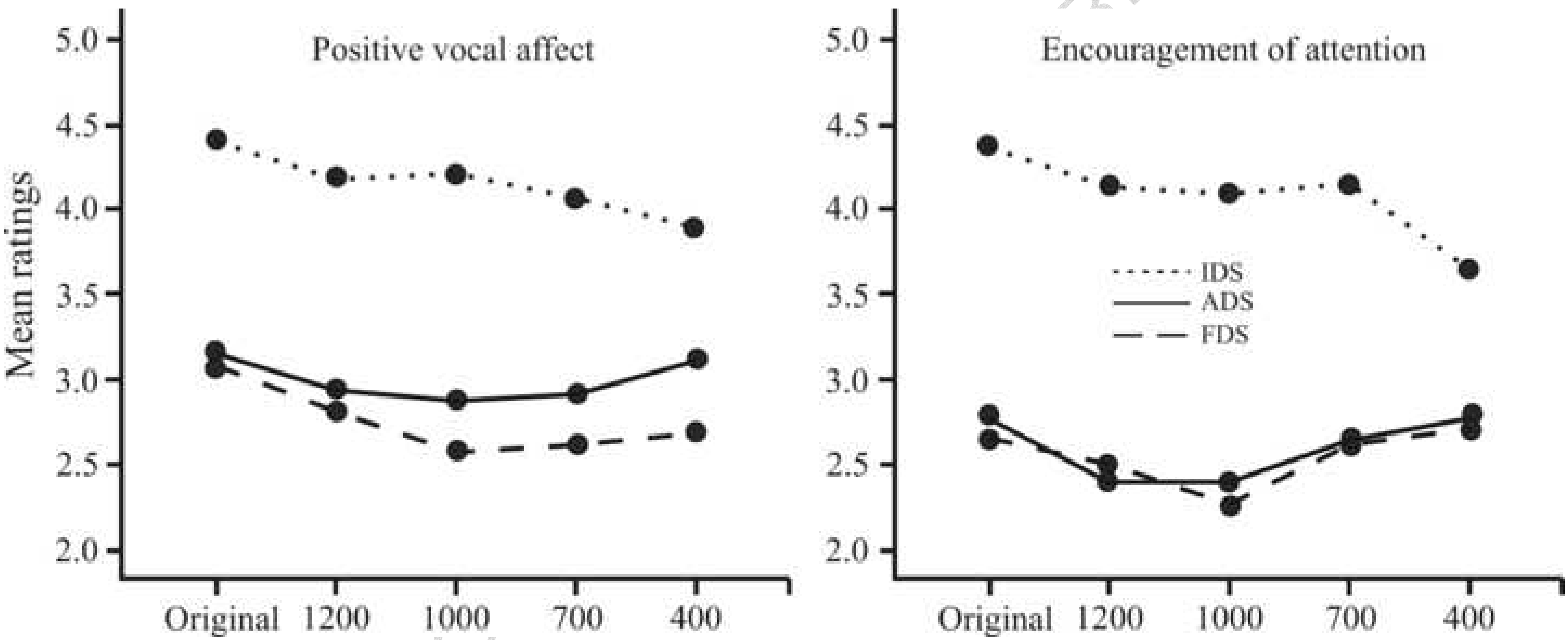
		Positive vocal affect			Encouragement of attention		
		Mean diff.	*F-value*	*Sig. level*	*Mean diff.*	*F-value*	*Sig. level*
Original	*IDS vs ADS*	1.246	93.333	$p < .001$	1.570	96.126	$p < .001$
	IDS vs FDS	1.312	129.441	$p < .001$	1.708	121.689	$p < .001$
	ADS vs FDS	Ns	Ns	Ns	Ns	Ns	Ns
1200	*IDS vs ADS*	1.241	90.927	$p < .001$	1.715	84.484	$p < .001$
	IDS vs FDS	1.363	89.918	$p < .001$	1.623	68.102	$p < .001$
	ADS vs FDS	Ns	Ns	Ns	Ns	Ns	Ns
1000	*IDS vs ADS*	1.333	102.19	$p < .001$	1.676	59.002	$p < .001$
	IDS vs FDS	1.633	108.924	$p < .001$	1.819	69.488	$p < .001$
	ADS vs FDS	-0.3	10.988	$p = .003$	Ns	Ns	Ns
700	*IDS vs ADS*	1.158	125.995	$p < .001$	1.496	177.724	$p < .001$
	IDS vs FDS	1.454	225.358	$p < .001$	1.521	134.683	$p < .001$
	ADS vs FDS	-0.296	13.844	$p = .001$	Ns	Ns	Ns
400	*IDS vs ADS*	0.776	40.825	$p < .001$	0.857	15.429	$p < .001$
	IDS vs FDS	1.205	72.318	$p < .001$	0.933	22.547	$p < .001$
	ADS vs FDS	-0.429	41.411	$p < .001$	Ns	Ns	Ns

ACCEPTED MANUSCRIPT

Figure captions

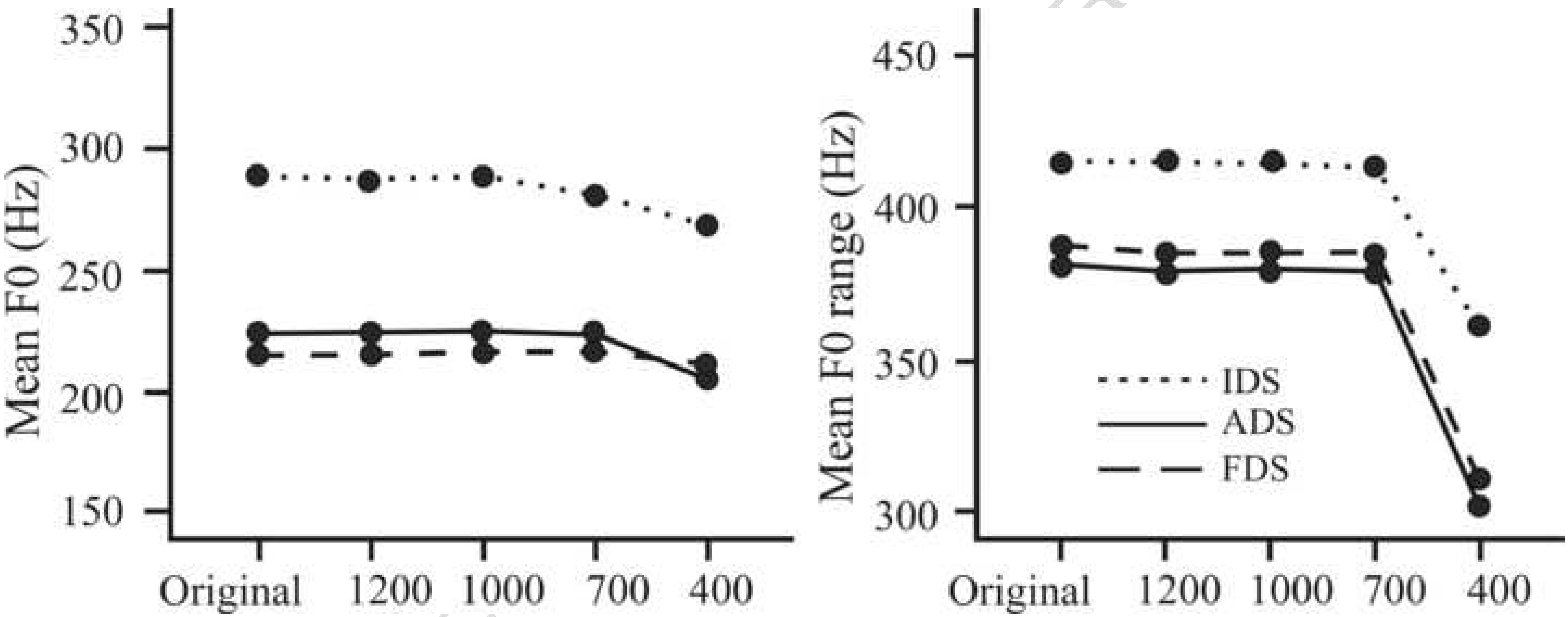
Fig. 1. Comparison of ratings for positive vocal affect (left) and encouragement of attention (right) for each filter presented for each of the three speech recipient groups. Note that of the two remaining scales, which are not presented here, the scale of negative vocal affect followed the trend of positive vocal affect, and the scales of comforting and soothing followed the trend of encouragement of attention.

Fig. 2. Comparison of mean F_0 (Hz; left) and F_0 range (Hz; right) for each of the three speech recipient groups across the five filters. Note that mean F_0 range is higher than in previous research (e.g. Fernald and Maizzie, 1991) as the analysis concentrated on the whole unmodified 30 seconds speech sample rather than on single utterances or words.



ACCEPTED MANUSCRIPT

Figure 1



ACCEPTED MANUSCRIPT

Figure 2