**Supplementary Methods**

**(i) Analysis of speech recordings**

We performed acoustic analyses of the speech sequences using a dedicated batch-processing script in PRAAT (version 6.0.04) with four distinct procedures. The first procedure of the script characterized the fundamental frequency (F0) and the intonation (F0 contour variation) of the speech sequence. In a first step, the F0 contour was extracted using the *To Pitch (cc)* command, and the following parameters were extracted: %voiced (percentage of the signal that is characterized by a detectable pitch, a measure of the proportion of spoken content), duration (total duration of the recording), mean F0, max F0, min F0 (respectively the mean, maximum and minimum F0 calculated over the duration of the signal) and F0CV (coefficient of variation of F0 over the duration of the signal). In a second step, two distinct smoothing algorithms (*Smooth* command in Praat) were performed on the pitch contour: the first allowed a relatively broad bandwidth (*Smooth* command parameter 25), to suppress very short-term frequency fluctuation while preserving minor intonation events, and the second only allowed a narrow bandwidth (*Smooth* command parameter 2), to only characterize strong F0 modulation (major intonation events). Inflection points were counted (as each change in the sign of the contour’s derivative) after each smoothing procedure, and divided by the total duration of the voiced segments in each recording, resulting in two distinct indexes of F0 variation (inflex25 and inflex2). A second procedure focused on the intensity contour and allowed the characterization of the variability of the speech sequence’s intensity by calculating intCV, the coefficient of variation of the intensity contour estimated using the *To intensity* command in PRAAT. A third procedure focused on the periodic quality of the signal and measured the harmonicity (harm, degree of acoustic periodicity, measured as the ratio of harmonics to noise in the signal and expressed in dB), an index of jitter (jitter, small fluctuation in periodicity measured as the average of ‘local’, ‘rap’ and ‘ppq5’ measures in PRAAT) and an index of shimmer (shimmer, small variation in amplitude between consecutive periods, measured as the average of ‘local’, ‘apq5’ and ‘apq11’ parameters in PRAAT). A final procedure characterized the first (lowest) five formant frequencies of the speech sequence. Formant frequencies were measured using the Linear Predictive Coding Burg algorithm in PRAAT (*To Formant (burg)* command), with a time step of 0.05 s, a maximum formant value of 5500 Hz, a window length of 0.1 s, and a pre-emphasis from 50 Hz. The mean formant frequencies (F1, F2, F3, F4, F5) were then averaged across the total duration of each speech sequence.

To test for differences in speech quality between the four recording conditions, we used linear mixed effect models with acoustic variables as dependent measures (fixed effect: recording condition –control, puppy, adult dog, old dog; random effect: speaker identity; package lme4, R version 3.1.2). P values were obtained with likelihood-ratio tests comparing the fit of full models with reduced models lacking the fixed effect. To compare between recording conditions, this analysis was followed by post-hoc (Tukey) multiple comparison tests (function glht in multcomp R package).

**(ii) Analysis of dogs’ behavioural reaction to playback**

The dog’s response was assessed using the 11 following behavioural measurements:

* latency to vocalize, i.e. time between the onset of played back signal and the first vocalization, calculated as a fraction (from 0 to 1) of the total duration of the experiment (value 1 if there was no vocalization during the whole playback experiment),
* latency to walk towards the speaker, i.e. time between the onset of the played back signal and the first dog’s walk, calculated as a fraction (from 0: immediate response to playback, to 1: no reaction during the experiment) of the total duration of the experiment,
* latency to look at the loudspeaker, i.e. time between the onset of the played back signal and the first look, calculated as a fraction (from 0: immediate response to playback, to 1: no reaction during the experiment) of the total duration of the experiment,
* number of looks at the speaker,
* number of vocalizations,
* number of looks at or walks towards the experimenter,
* number of stays at less than 1 meter from the loudspeaker,
* number of walks towards the loudspeaker,
* duration of stays at less than 1 meter from the loudspeaker,
* total duration of looks toward the loudspeaker,
* duration of tilted head position towards the loudspeaker.

Instead of separately analyzing the dependent behavioural measures, we performed a principal component analysis (PCA) and retained a single composite score (PC1), separately for each of the two experiments [28]. By using a PCA, we approached a Gaussian distribution and built an integrated measure of the behavioural response. The effect of factors on this integrated measure was then assessed using a GLM approach (R software; dependent variable = integrated measure of the dog’s behavioural reaction PC1; fixed factors = speech quality, dog’s age, order of the playback test, interactions between speech quality, age and order for the first experiment, type of dog-directed speech and playback order for the second experiment; random factors = dog’s identity, and dog origin –shelter *vs* family- for the first experiment). Speech quality factor had two levels: puppy-directed speech and human-directed speech. Dog’s age factor had also two levels: puppy and adult. In order to investigate the effect of specific acoustic features characterizing speech sequences on dog behaviour, we built models with acoustic features, dog age, their interactions and playback order as fixed factors. All models had a Maximal Random Effects structure, with random slopes within the dog’s identity for each fixed effect [29]. P values were obtained with likelihood-ratio tests comparing the fit of full models with reduced models lacking each fixed effect. These analyses were followed by Tukey post-hoc tests.